University of Alberta

DESIGN AND ENGINEERING OF STABLE HAIRPIN PEPTIDE SCAFFOLDS FOR MOLECULAR RECOGNITION

by

Terrance Zihao Cheng

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry

Edmonton, Alberta
Fall 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-46295-9
Our file Notre référence
ISBN: 978-0-494-46295-9

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

To my parents and grandparents. Without your unconditional love and unwavering support all the way, I wouldn't be the first Ph.D. in our family. I love you.

This work is also dedicated to my lovely fiancée, June, who has accompanied with me throughout the course of this thesis. You are the one who has made my Ph.D. journey colorful!

ABSTRACT

Binding proteins suitable for expression and high affinity molecular recognition in the cytoplasm or nucleus of live cells have numerous applications in the biological sciences. In an effort to add a new minimal motif to the growing repertoire of validated non-immunoglobulin binding proteins, we have undertaken the further development of a generic protein scaffold based on a "tryptophan zipper" (trpzip) class of β-hairpin peptides that can fold efficiently in the cytoplasm. A method has been developed, based on the measurement of fluorescence resonance energy transfer (FRET) between a genetically fused cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), which allows the structural stability of recombinant β-hairpin peptides to be rapidly assessed both *in vitro* and *in vivo*. The results suggest that even the most highly folded previously reported member of the trpzip class (the 16mer **HP5W4**) was not qualified to be a peptide scaffold because of its poor structural stability upon introduction of destabilizing substitutions [1]. To address this issue, we used a fluorescence-based live cell screening system to identify extended trpzip variants with additional stabilizing interactions. Our results support the conclusion that two variants, a 20mer and a 24mer peptide, are structured even in the presence of randomized amino-acid substitutions or insertions [2, 3]. This superb stability makes them suitable to be converted into molecular recognition domains. Randomized amino acid residues were genetically introduced. The peptide libraries were displayed on the surface of filamentous phage M13 as a low copy number pIII fusion. Panning of the resulting libraries against immobilized streptavidin resulted in a dramatic enrichment of phage displaying peptide sequences with submicromolar affinities for streptavidin. Mutating key structurally important residues of the peptide resulted in a 3 order of magnitude drop in the binding affinity, illustrating the critical importance of the scaffold structure for high affinity molecular recognition.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Robert E. Campbell, for his guidance during my study at the University of Alberta. I have always felt very fortunate to work with such an ingenious leader who is always available and willing to help. Without his support and continual encouragement, my doctoral training would not be so smooth and rewarding. I also wish to thank all of the Campbell group members, both past and present, for their generous help. In particular, I want to thank Ahmed Ibraheem, who did extensive proof-reading for me.

I also would like to sincerely thank my supervisory and examination committee, for their participation in my candidacy and final oral examination, precious advice regarding my research, as well as my thesis adjustment. I also owe my thanks to Wayne Moffat, Dr. Sandra Marcus, Dr. Eric Pelletier, Mark Miskolzie and Svetlana Sapelnikova, for their technical assistance. I thank the Cold Spring Harbor Laboratory for offering the phage-display course, and Dr. Carlos F. Barbas for sharing his phage-display vectors.

I cannot end without thanking my parents, grandparents and fiancée. I owe my loving thanks to them. Without their love and understanding, it would have been impossible for me to finish this work.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION 1

PREAMBLE 2

1.1 FLUORESCENT PROTEINS 3

1.1.1 Green fluorescent protein 4

1.1.2 Fluorescent proteins of various colors 7

1.2 FLUORESCENCE RESONANCE ENERGY TRANSFTER (FRET) 9

1.2.1 FRET efficiency, definition and calculation 10

1.2.2 FRET in fluorescent proteins 12

1.2.3 FRET in monitoring protein structure and conformational change 14

1.2.3.1 Intermolecular FRET-based sensors 15

1.2.3.2 Intramolecular FRET-based sensors 17

1.3 NOVEL PROTEINS FOR MOLECULAR RECOGNITION 18

1.3.1 Immunoglobulin or non-immunoglobulin? 19

1.3.2 Novel non-immunoglobulin binding proteins 21

1.3.2.1 Immunoglobulin-like domains 22

1.3.2.2 Non-immunoglobulin scaffolds with α-helix frameworks 23

1.3.2.3 Non-immunoglobulin scaffolds with β-sheet frameworks 26

1.3.2.4 Non-immunoglobulin scaffolds with combined structures 27

1.3.3 Phage display libraries and their screening 28

1.4 SUMMARY.....36

CHAPTER 2: ASSESSING THE STRUCTURAL STABILITY OF DESIGNED β-HAIRPIN PEPTIDES IN THE CYTOPLASM OF LIVE CELLS.....37

2.1 INTRODUCTION.....38

2.2 MATERIAL AND METHODS.....42

2.2.1 General procedures and materials.....42

2.2.2 Construction of CFP-peptide-YFP expression vectors.....43

2.2.3 Construction of CFP-**tz1**-peptide expression vector.....45

2.2.4 Protein and peptide expression and purification.....45

2.2.5 Spectroscopy.....46

2.2.6 Description of the bacterial colony imaging system.....47

2.3 RESULTS AND DISCUSSION.....48

2.3.1 Construction of CFP-peptide-YFP expression vectors.....48

2.3.2 *In vitro* characterization of **tz** series proteins.....49

2.3.3 *In vivo* characterization of **tz** series proteins.....53

2.4 CONCLUSION.....57

CHAPTER 3: A RATIONAL APPROACH TO DEVELOP AN EXTENDED 20-MER β-HAIRPIN.....59

3.1 INTROUDCTION.....60

3.2 MATERIAL AND METHODS.....63

3.2.1 General procedures and materials.....63

3.2.2 Constructions of CFP-**xtz2**-YFP and CFP-**xtz2**-peptide expression vectors....63

3.2.3 Protein and peptide purification....64

3.2.4 Spectroscopy....64

3.3 RESULTS AND DISCUSSION....64

3.3.1 *In vitro* characterization of **xtz2** series proteins....64

3.3.2 *In vivo* characterization of **xtz2** series proteins....67

3.3.3 Characterization of the isolated peptide portion of **xtz2**....69

3.4 CONCLUSION....70

CHAPTER 4: *IN VIVO* SCREENING IDENTIFIES A HIGHLY FOLDED β-HAIRPIN PEPITDE WITH A STRUCTURED EXTENSION....72

4.1 INTRODUCTION....73

4.2 MATERIAL AND METHODS....74

4.2.1 General procedures and materials....74

4.2.2 Library construction and screening....74

4.2.3 Protein and peptide purification....75

4.2.4 Spectroscopy....76

4.2.5 Protein expression and imaging in HeLa cells....76

4.2.6 NMR spectroscopy and structure calculation....77

4.3 RESULTS AND DISCUSSION....78

4.3.1 An "irrational" approach to develop an extended and highly folded 20-mer β-Hairpin....78

4.3.2 *In vitro* characterization of **xtz1** series proteins 80

4.3.3 Characterization of the isolated peptide portion 82

4.3.4 **xtz1** structure elucidated by NMR 83

4.3.5 Assessing the conformational stability in mammalian cells 88

4.3.6 Beyond the 20-mer: efforts directed toward the development of 24-mer and 28-mer β-hairpins 90

4.4 CONCLUSION 92

CHAPTER 5: ENGINEERED TRYPTOPHAN ZIPPER-TYPE PEPTIDES AS MOLECULAR RECOGNITION SCAFFOLDS 94

5.1 INTRODUCTION 95

5.2 MATERIAL AND METHODS 99

5.2.1 General procedures and materials 99

5.2.2 Library construction and screening 99

5.2.3 Protein and peptide purification 100

5.2.4 Spectroscopy 100

5.2.5 Phage displayed peptide library construction and screening 100

5.2.6 Surface plasmon resonance 102

5.3 RESULTS AND DISCUSSION 102

5.3.1 *In vivo* screening to identify a 24-mer hairpin peptide 102

5.3.2 Characterization of **xxtz1** series proteins 104

5.3.3 Characterization of **xxtz1**-peptide 106

5.3.4 Phage display libraries construction and screening 109

5.3.5 Characterization of streptavidin binding proteins.....113

5.4 CONCLUSION.....114

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS.....116

6.1 FRET APPROACH IN PROTEIN ENGINEERING.....117

6.2 TRPZIP VARIANTS SCAFFOLDS.....118

APPENDIX.....141

XTZ1-PEPTIDE NMR AND STRUCTURE DATA.....141

LIST OF TABLES

Table 1 Sequences of **tz** series peptides flanked by CFP and YFP 48

Table 2 FRET efficiencies, rates of proteolysis, and exciton strengths for proteins **tz1** through **tz6**.....50

Table 3 Sequences of **xtz** series peptides flanked by CFP and YFP 63

Table 4 FRET efficiencies and rates of proteolysis for proteins **tz1** and **xtz2** through **xtz5**..... 65

Table 5 Peptide sequences identified by library screening in **Chapter 4** 79

Table 6 Summary of proteins and peptides compared in **Chapter 4** 80

Table 7 FRET efficiencies and rates of proteolysis for proteins **tz1**, **xtz1** through **xtz7**..... 81

Table 8 Sequences of the 24-mer and 28-mer peptide libraries 91

Table 9 Peptide sequences identified by library screening in **Chapter 5** 103

Table 10 Summary of proteins and peptides compared in **Chapter 5** 104

Table 11 **tz1**, **xtz1** and **xxtz** series proteins FRET comparison 105

Table 12 Phage display libraries.....110

Table 13 Peptide sequences obtained from panning against streptavidin 112

Table A1 Chemical shifts for **xtz1**-peptide 142

Table A2 Statistics for the 20 model NMR solution structures of **xtz1**-peptide 144

LIST OF FIGURES

Figure 1 A ribbon representation of the three-dimensional structure of avGFP 5

Figure 2 Mechanism of avGFP chromophore formation.....6

Figure 3 Schematic representation of chimera protein production.....7

Figure 4 Chromophore structures of BFP, CFP, EGFP, YFP and DsRed.....8

Figure 5 Jablonski diagram of FRET.....9

Figure 6 The dependence of the FRET efficiency on interchromophore distance 11

Figure 7 General design of FRET-based FP probes 16

Figure 8 Immunoglobulin G structure 20

Figure 9 Structures of immunoglobulin-like domains.....23

Figure 10 Non-immunoglobulin scaffolds currently used for molecular recognition.....25

Figure 11 Schematic representation of the filamentous baceteriophage particle 30

Figure 12 A schematic representation of the phage DNA replication, protein synthesis and protein location in a Ff phage cycle.....32

Figure 13 Schematic representation of foreign protein displayed on the surface of filamentous phage.....33

Figure 14 Screening a phage displayed library.....35

Figure 15 Schematic illustration of the FRET efficiency and rate of trypsin digestion. .41

Figure 16 The corrected and normalized emission spectra of **tz1** and **tz6**. 49

Figure 17 Ratio of YFP to CFP for purified proteins **tz1** through **tz6** upon treatment with trypsin..... 51

Figure 18 The far-UV CD spectra of identical concentrations of **tz1** and **tz6**.....52

Figure 19 Difference CD spectra obtained by subtracting the CD spectrum of trypsinized **tz6** from each of **tz1** through **tz6** 53

Figure 20 The YFP/CFP ratio of proteins **tz1** through **tz6** in live cells.....54

Figure 21 Relationship between the YFP/CFP peak ratio and the protein expression level.....55

Figure 22 Concentration dependent fluorescent spectra.....56

Figure 23 Representations of the β-hairpin peptide in the context of the CFP/YFP fusion protein.....62

Figure 24 Ratio of YFP to CFP for purified proteins **xtz2** through **xtz5** upon treatment with trypsin.....66

Figure 25 Difference CD spectra obtained by subtracting the CD spectrum of trypsinized **xtz3** from each of intact **xtz2**, **xtz4**, and **xtz5**.....67

Figure 26 The YFP/CFP ratio of proteins **xtz2** through **xtz5** in live cells.....68

Figure 27 CD spectra of identical concentration of **HP5W4**, **xtz2**-peptide in buffer and **xtz2**-peptide in 50% TFE.....70

Figure 28 CD spectra and melting curves of **HP5W4**, **xtz1**-peptide and **xtz2**-peptide...83

Figure 29 Stereoview of the **xtz1**-peptide NMR structure 84

Figure 30 Stereoview of a superposition of the NMR structure ensembles of **HP7** and **xtz1**-peptide.....85

Figure 31 Chemical shift deviation (ĆSD) analysis of Hα, HN, and side chain hydrogen atoms of **xtz1**-peptide 86

Figure 32 A close-up look at the "extended" residues.....87

Figure 33 Fluorescence imaging of the **tz1**, **xtz1**, **xtz2**, **xtz3** FRET constructs and the FRET efficiencies comparisons *in vitro* and in HeLa cells 90

Figure 34 Schematic representation of "stem-loop" type structure based on SKVILFE repeat.....96

Figure 35 Schematic illustration of the **xxtz1**-peptide engineering and the phage display libraries used in **Chapter 5**.....98

Figure 36 CD spectra and melting curves of **HP5W4**, **xtz1**-peptide, **xxtz1**-peptide and **xxtz1(loop)**-peptide 108

Figure 37 Phage output in each round of selection.....111

Figure 38 Sensorgrams of the binding of **xxtz1-HPQ1** fusion proteins to a streptavidin coating surface.....114

LIST OF ABBREVIATIONS

ε	Extinction coefficient
ϕ	Fluorescence quantum yield
κ^2	Förster orientation factor
μM	Micromolar (10^{-6} mole per liter)
2D	Two-dimensional
Å	Angstrom (10^{-10}m)
A	FRET acceptor
avGFP	*Aequorea victoria* green fluorescent protein
BFP	Blue fluorescent protein
CD	Circular dichroism
cDNA	Complementary DNA - DNA copied from a mature mRNA template using reverse transcriptase
CDR	Complementarity determining region
CFP	Cyan fluorescent protein
C_H/ C_L	Heavy/light domain of the constant region of an antibody
CSD	Chemical shift deviation
D	FRET donor
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double stranded DNA
DsRed	*Discosoma sp.* red fluorescent protein
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
E	FRET efficiency
E. coli	Escherichia coli

EBFP	Enhanced blue fluorescent protein
ECFP	Enhanced cyan fluorescent protein
EGFP	Enhanced green fluorescent protein
EtF	Edge-to-face
Fab	Fragment antigen binding, region of an antibody
FACS	Fluorescence activated cell sorter/sorting
FBS	Fetal (or Foetal) bovine serum
Fc	Fragment crystallizable region, constant region of an antibody
FP	Fluorescent protein
FRET	Fluorescence (or Förster) resonance energy transfer
Fv	Variable fragment of an antibody
GB1p	B1 domain of the IgG binding protein in *Streptococcus*
gDQF-COSY	Gradient enhanced double quantum filter correlation spectroscopy
GFP	Green fluorescent protein
HBSS	Hanks balanced salt solution
HeLa	Immortal cell line derived from Henrietta Lacks cervical cancer cells
HPLC	High performance liquid chromatography
Ig	Immunoglobulin
IgG	Immunoglobulin of the G class/isotype
$J(\lambda)$	FRET donor and acceptor spectra overlap integral
K_d	Dissociation constant
kD	KiloDalton
LB	Luria-Bertini medium
M	Molar (mole per liter)
MALDI	Matrix-assisted laser desorption ionization
mM	Millimolar (10^{-3} mole per liter)
ms	Millisecond (10^{-3} second)

mTFP	Monomeric teal fluorescent protein
n	Refractive index
NAR	New antigen receptor
Ni-NTA	Ni^{2+}-nitrilotriacetate
nm	Nanometer (10^{-9} m)
NOESY	Nuclear overhauser enhancement spectroscopy
OptiMEM	Modified Eagle's minimum essential medium, reduced serum
pI	Ff filamentous phage *gene I* protein
pIII	Ff filamentous phage *gene III* protein
pIV	Ff filamentous phage *gene IV* protein
pV	Ff filamentous phage *gene V* protein
pVI	Ff filamentous phage *gene VI* protein
pVII	Ff filamentous phage *gene VII* protein
pVIII	Ff filamentous phage *gene VIII* protein
pIX	Ff filamentous phage *gene IX* protein
pX	Ff filamentous phage *gene X* protein
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data base
PEG	Polyethylene glycol
Q_D	Quantum yield of the donor in the absence of the acceptor
r	Distance between donor and acceptor of FRET pair
R_0	Förster radius
rcf	relative centrifugal force
RET	Resonance energy transfer
RFP	Red fluorescent protein
rpm	Revolutions per minute

SB	Super broth medium
scFv	Single chain variable fragment
SPR	Surface plasmon resonance
ssDNA	Single-stranded DNA
TFE	Trifluoroethanol
TLC	Thin layer chromatography
TOCSY	Total correlation spectroscopy
TOF	Time-of-flight
Trpzip	Tryptophan zipper
tz	Trpzip (tryptophan zipper)
UV	Ultraviolet
V_H	Heavy chain of an antibody variable domain
V_L	Light chain of an antibody variable domain
VhH	Heavy chain antibody VH
XRD	X-ray diffraction
xtz	Extended trpzip
YFP	Yellow fluorescent protein
YPet	Yellow fluorescent protein for energy transfer

CHAPTER 1:

INTRODUCTION

PREAMBLE

Protein engineering is the science of designing, engineering and selecting useful proteins with desired properties for different applications. One of the landmark achievements in this field is the engineering of a useful fluorescent protein (FP) for live cell imaging [4]. Starting from the first known FP, *Aequorea victoria* green fluorescent protein (avGFP), numerous FPs have been engineered with properties such as altered color, increased brightness, and improved photostability. The availability of these engineered FPs is considered by many to be a biotechnological breakthrough, as they have facilitated research in almost all areas of biochemistry, cell biology, and physiology. An overview of avGFP and its variants and homologues will be presented in the following section. This will be followed by a discussion of FRET and its application in monitoring protein structure and conformational changes.

Engineering novel protein scaffolds for specific molecular recognition is another milestone in protein engineering. Antibodies are unrivalled in terms of versatility and practicality for molecular recognition of almost any target molecule. They can bind to a wide variety of target molecules with high affinity and high specificity. Although they are superb in most of the cases, they are not optimal in some applications such as molecular recognition in the reducing environment of the cytoplasm. The third section in this introduction chapter will provide an overview of novel proteins that have been engineered for molecular recognition. In addition, phage display and the *in vitro* methodology for screening phage libraries will also be discussed.

Chapter 2 addresses my hypothesis that fluorescence resonance energy transfer (FRET) is an efficient approach to rapidly reveal genetically fused peptide structures both *in vitro* and *in vivo*. Both **Chapter 3** and **4** address my hypothesis that rational and irrational protein engineering strategy would be able to increase my target peptide stability in tolerating destabilizing amino-acid substitution. **Chapter 5** will show the application of the engineered peptide as a generic scaffold for molecular recognition, followed by a conclusion chapter in **Chapter 6**.

1.1 FLUORESCENT PROTEINS

In recent years fluorescence-based detection has emerged as a key methodology in chemistry and biology because of its high sensitivity and simple instrumental requirements. For example, fluorescence is used extensively in organic chemistry to monitor the separation of organic compounds on thin layer chromatography (TLC). In analytical chemistry, fluorescence can be used in high performance liquid chromatography (HPLC) detectors for highly sensitive detection of trace analytes. The most remarkable growth in the use of fluorescence has been witnessed in biological sciences for the past three decades. Fluorescence has become an important means in a variety of biotechnological fields such as DNA sequencing, immunohistochemistry, and fluorescent-activated cell sorting (FACS), among others. Fluorescence is also widely applied in live cell imaging to monitor protein localization, organelle structures, enzyme activities, small molecule concentrations, and protein-protein interactions.

By definition, a fluorescent molecule can absorb a photon at a particular energy and then re-emit a photon at lower energy (longer wavelength). Analogous to chromophores that absorb light (and may or may not be fluorescent), fluorescent molecules or fluorescing portions of a larger molecule are called fluorophores. Generally, biological fluorophores can be divided into two groups – intrinsic and extrinsic. Intrinsic fluorophores are fluorophores that occur naturally like aromatic amino acids and flavins. Extrinsic fluorophores, on the other hand, are typically synthetic dyes or biochemicals with specific spectral properties that are used to label other molecules.

In proteins, tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe) are fluorescent when excited with UV light, so they are classified as intrinsic fluorophores. Tryphophan fluorescence has proven exceptionally useful for biophysical characterization of protein dynamics and ligand binding. However, the intrinsic fluorescence of proteins is not particularly useful in imaging applications because it is ubiquitous in living cells. Furthermore, excitation of intrinsic protein fluorescence requires high energy UV light that is damaging to DNA and the resulting fluorescence is limited by relatively low quantum yields and extinction coefficients. Extrinsic fluorophores, on the other hand,

usually have excellent fluorescent properties, such as high brightness, good photostability, and are available in a wide selection of colors. Some well known dyes, fluorescein and rhodamine are good examples. However, extrinsic fluorophores are not always user-friendly when used to label proteins and tedious processes procedures may be involved. Furthermore, the labelled protein is structurally heterogenous since each protein molecule could react with a variable number of dye molecules. Genetic labelling with FPs provides an alternative method of introducing fluorescence into proteins and addresses some of the limitations of both intrinsic and extrinsic fluorophores. Arguably, FPs are neither an intrinsic nor an extrinsic fluorophore but represent a class unto themselves.

1.1.1 Green fluorescent protein

FPs are now known to exist in various coelenterates (aquatic animals which include the jellyfish, coral, sea anemone and hydroids). The first FPs to be identified was avGFP from *Aequorea* jellyfish. The avGFP had, of course, existed for millions of years in *Aequorea* jellyfish, but humans were not aware of its existence until it was discovered and isolated in 1962 by Shimomura *et al.,* who squeezed the proteins from the bioluminescent organs of thousands of jellyfish [5]. During the protein purification and separation process, Shimomura noticed a protein "...giving solutions that look slightly greenish in sunlight though only yellowish under tungsten lights, and exhibiting a very bright, greenish fluorescence in the ultraviolet of a Mineralite [a handheld UV lamp]..." This protein, which is now known as avGFP, was characterized and its emission spectrum was recorded in the same year [5]. The avGFP is composed of 238 amino acids with a molecular weight of 26.9 kD. The avGFP and its isoforms sequences were published in the 1990s [4, 6]. Although first crystallized in 1974 [7], the avGFP structure was not solved until 1996 [8, 9]. The avGFP has a unique β-barrel structure (**Figure 1**) made up of eleven β-strands with an α-helix running through the center.

Figure 1 A ribbon representation of the three-dimensional structure of avGFP (protein data base (PDB) 1EMA) [8]. Graphical representations were prepared by YASARA.

In the middle of the α-helix lies the chromophore, *p*-hydroxybenzylideneimidazoline, which is formed by the autocatalytic cyclization of residues Ser65, Tyr66, and Gly67 in the α-helix. A mechanism for the chromophore formation has been proposed [10-12], and can be described as a three-step process. The cyclization is initiated by the nucleophilic attack of Gly67 backbone nitrogen atom on the Ser65 backbone carbonyl carbon, creating the imidazolone ring. Following elimination of a molecule of water, a further conjugation step involving Tyr66 oxidation occurs (**Figure 2**). Because the chromophore is located at the approximate β-barrel geometric centre, it has been referred as a "light in the can".

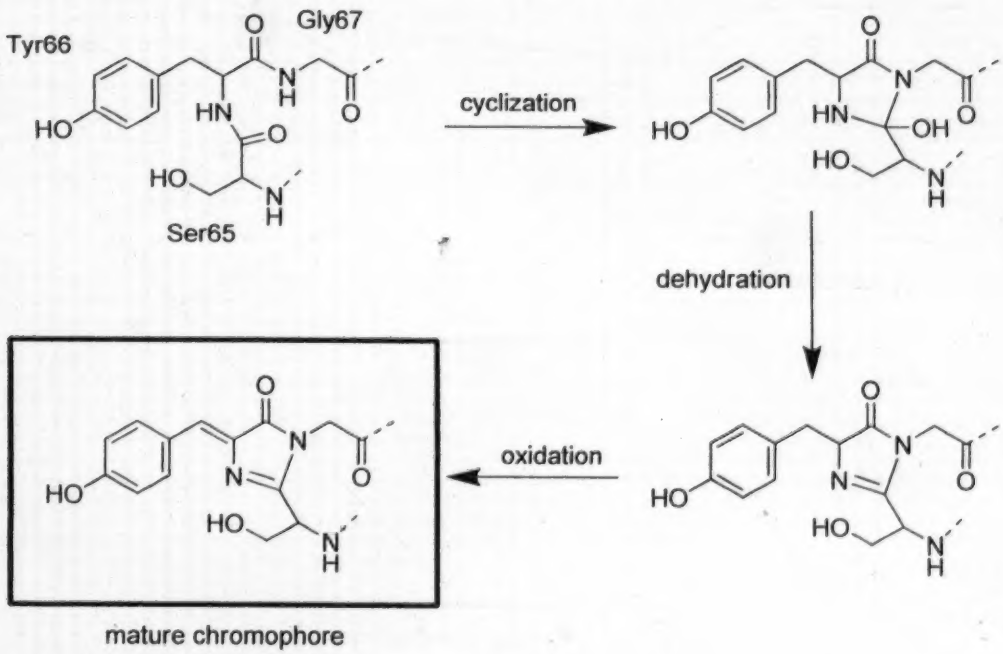


Figure 2 Mechanism of avGFP chromophore formation [11, 12].

In the 1992, the avGFP nucleotide sequences was cloned from *Aequorea victoria* [6]. Soon after, it was shown that the wild-type avGFP can be expressed in a functional form from its cDNA in both prokaryotic (*E. coli*) and eukaryotic (*C. elegans*) cells [13]. Since then, numerous efforts have been invested into wild-type avGFP optimization by different mutational strategies, such as error-prone PCR [12, 14-16], deliberate site-directed mutations [8, 9, 17, 18], saturated mutagenesis [18, 19], DNA shuffling [20], and heteroduplex recombination [21]. Variants of avGFP that have increased folding efficiency at 37°C, increased brightness, simplified excitation spectrum, reduced sensitivity to pH and halides, reduced tendency to dimerize, and different colors have been engineered [4]. Application of avGFP has expanded from being used as a gene-expression reporter [13] to numerous applications in different fields. The most successful avGFP application so far is as a genetic fusion partner to different proteins. With modern molecular biology techniques, the gene encoding FP can be easily ligated with genes coding any endogenous protein (**Figure 3**). The resulting chimeric protein can be

expressed in live cells in culture or in different organisms. By combining directed and random mutagenesis, avGFP has been engineered to function as a pH sensor [22]. Different avGFP fusion chimeric proteins have been developed as active indicators of Ca^{2+} ions [23, 24] and protease activity [25].

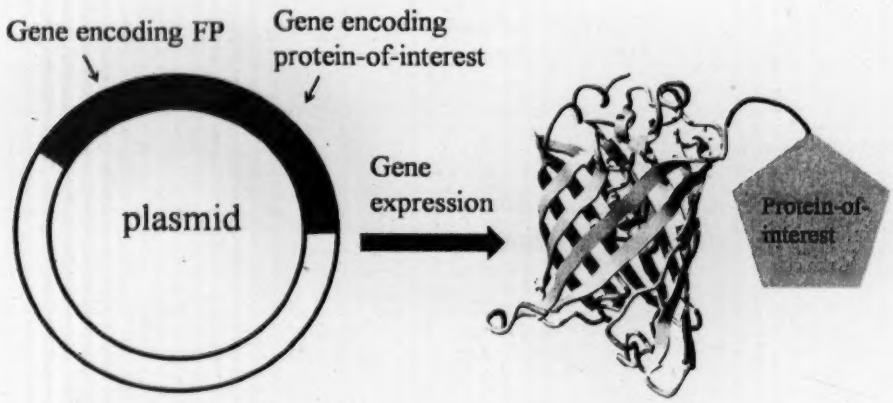


Figure 3 Schematic representation of chimera protein production.

1.1.2 Fluorescent proteins of various colors

Almost immediately after avGFP was cloned in 1992, it became an indispensable tool to biochemists. However, avGFP did have the obvious limitation of only being available in one color originally. To address this limitation, there has been a continuing effort to develop more color variants of FPs to add to the "toolkit" of useful variants. Different mutagenesis strategies have been applied to avGFP, which generally fall into two categories: creating chromophore variants and changing the chromophore environment. The former approach was used to generate the blue fluorescent protein (BFP) [26] and the cyan fluorescent protein (CFP) [12]. The latter was used to create enhanced green fluorescent protein (EGFP) [26, 27], Sapphire [26], mKalama1 [28], and yellow fluorescent protein (YFP) [26].

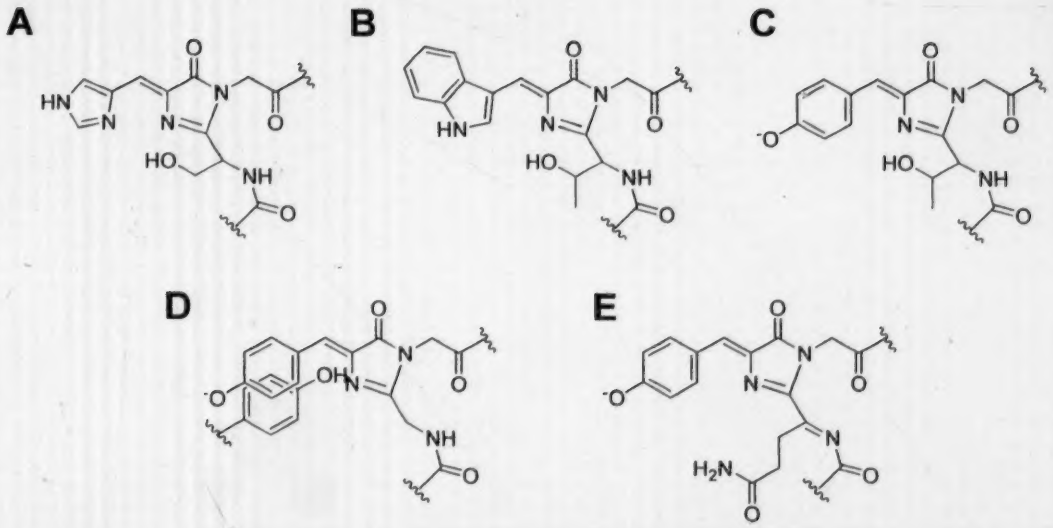


Figure 4 Chromophore structures of (**A**) BFP, (**B**) CFP, (**C**) EGFP, (**D**) YFP, and (**E**) DsRed.

BFP was created by replacing the amino acid residue Tyr66 with a His (Tyr66His), which results in blue-shifted excitation and emission [26] (**Figure 4A**). CFP, on the other hand, was obtained by replacing Tyr66 with a Trp (Tyr66Trp) [12] (**Figure 4B**). Wild-type avGFP chromophore exists in two states: anionic and neutral. By introduction of amino acid substitutions that change chromophore environment, EGFP [26, 27] (**Figure 4C**) and Sapphire [26] were obtained by suppressing anionic or protonated chromophore, respectively. YFP was engineered by introduction of the Tyr (Thr203Tyr) amino acid substitution [26] (**Figure 4D**). An X-ray crystal structure confirmed that the aromatic side chain of Tyr stacks next to the chromophore [26]. It has been proposed that the π - π interactions reduce the excited-state energy level, resulting in red-shifted excitation and emission [29].

Besides jellyfish, corals provide another source of FPs with more choices of colors from cyan to red. The first red fluorescent protein (RFP) was cloned from the coral *Discosoma* (this RFP is known as DsRed) [30]. Extensive study of DsRed revealed that the red-shifted emission spectrum results from extended conjugation of the chromophore [31-33]

(**Figure 4E**). FPs of different colors not only make labeling of multiple proteins in a single cell possible, but also they pave the way for conducting FRET experiments *via* FP pairs.

1.2 FLUORESCENCE RESONANCE ENERGY TRANSFTER (FRET)

Fluorescent resonance energy transfer or Förster resonance energy transfer (FRET), was theoretically described by Förster in 1948 [34]. FRET results from a dipole-dipole interaction between a donor (D) molecule in the excited state and an acceptor (A) molecule in the ground state. FRET only occurs when the donor and the acceptor are in close proximity, and when the donor emission spectrum overlaps with the acceptor absorption spectrum. Instead of involving collision or generating heat, the excited donor decays radiationlessly by transferring energy to the ground state acceptor, which can be either fluorescent or nonfluorescent (**Figure 5**).

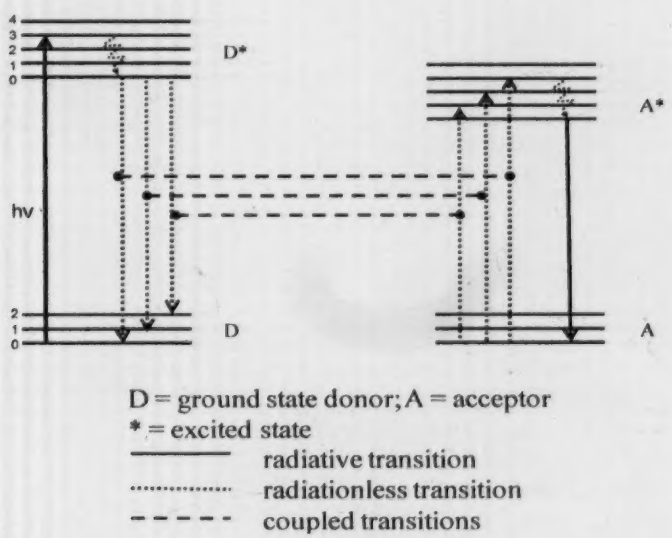


Figure 5 Jablonski diagram of FRET (modified from Moens *et al.* [35]). It shows the energy arrangement for a FRET donor-acceptor pair.

Because it is the excitation energy rather than fluorescence energy that is transferred, it was long argued that the most appropriate acronym should be RET (resonance energy transfer) rather than FRET [36]. Nevertheless, the letter "F" represents fluorescence which is a valuable descriptive term of RET; and "F" also represents the great contribution from Förster. For these reasons, the abbreviation FRET will be used in this thesis.

1.2.1 FRET efficiency, definition and calculation

The FRET efficiency (E) can be obtained experimentally by:

$$E = 1 - \frac{F_{DA}}{F_D} \qquad \textbf{Equation 1.1}$$

where F_{DA} and F_D are the donor fluorescent intensity in the presence (F_{DA}) and absence (F_D) of acceptor. FRET efficiency can be theoretically defined using the equation:

$$E = \frac{R_0^6}{R_0^6 + r^6} \qquad \textbf{Equation 1.2}$$

where R_0 is the Förster distance (see below) and r is the distance between the donor and acceptor. From this equation it is obvious that FRET efficiency is strongly (6^{th} power) distance dependent. The graphic relationship is illustrated in **Figure 6**. FRET efficiency is most strongly dependent on distance when $r = R_0$. It should be noticed that it is not practical to measure FRET efficiency outside the range of r less than $0.5R_0$ and r greater than $2R_0$, as the slope of the curve becomes ever less responsive to the distance change.

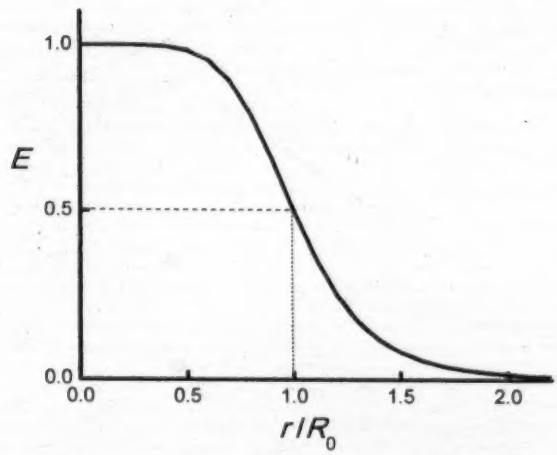


Figure 6 The dependence of the FRET efficiency (E) on interchromophore distance.

The Förster distance R_0 in **Equation 1.2** was originally defined as the distance between the donor and the acceptor at which the energy transfer is at 50% efficiency. R_0 can be calculated by:

$$R_0^6 = 8.79 \times 10^{23} [\kappa^2 n^{-4} Q_D J(\lambda)] \text{ (in Å}^6\text{)} \qquad \textbf{Equation 1.3}$$

The κ^2 term is the Förster orientation factor, which can theoretically range from 0 to 4; n is the refractive index, which is usually assumed to be 1.4 for biomolecules in aqueous solution. The Q_D term is the quantum yield of the donor in the absence of the acceptor. $J(\lambda)$ is the overlap integral, representing the degree of spectral overlap between the donor emission and the acceptor absorption. All of the parameters necessary for calculation of R_0 can be determined experimentally except the κ^2. A κ^2 value of 0 represents the case in which one of the transition dipoles is perpendicular to the other, while 4 represents the case in which the transition dipoles are parallel.

As a result, it is clear that a FRET change during experiments carried out under the same condition is arisen from the change of one or both of the two variants, distance and/or orientation [37]. This relationship can be described mathematically using **Equation 1.4**, where κ^2 is the dipole-dipole orientation factor, and r is the interchromophore distance,

and C is a constant. This equation reveals that a FRET change (ΔE) is determined by the interchromophore distances ($r_{initial}$ and r_{final}) and the orientation factors ($\kappa^2_{initial}$ and κ^2_{final}) in the initial and final states, respectively.

$$\Delta E = \frac{1}{1+\left(\frac{r^6_{final}}{C \times \kappa^2_{final}}\right)} - \frac{1}{1+\left(\frac{r^6_{initial}}{C \times \kappa^2_{initial}}\right)} \qquad \textbf{Equation 1.4}$$

The two extreme values (0 and 4) of κ^2 require the donor and acceptor to be exactly and rigidly oriented in space in particular orientations. Such a scenario is not typically found in FRET experiments. Rather, the κ^2 value is generally assumed to be equal to 2/3, provided that the probes are in fast isotropic motion [35]. This assumption was confirmed by comparing experimental FRET data with X-ray diffraction (XRD) data [37]. Since the sixth power root of distance is taken in calculating the FRET efficiency, the distance dependence in FRET efficiency calculation generally outweighs the uncertainty arising from the orientation factor. Therefore, FRET is a reliable "spectroscopic ruler" in measuring structural changes between two fluorophores.

1.2.2 FRET in fluorescent proteins

Today, thousands of organic dyes are available for different application purposes. Many of them could be potential FRET pairs. But what makes a good FRET pair? Generally, a good FRET pair should have a large R_0 that means greater FRET efficiency at a greater interchromophore distance. As shown in **Equation 1.3**, a larger R_0 can be obtained by using FRET pairs with better overlap between donor emission and acceptor excitation profiles (larger $J(\lambda)$). A brighter donor (larger Q_D) will also result in a larger R_0. The refractive index (n) is effectively constant (about 1.4 for aqueous solutions) for any particular FRET experiment. The remaining parameter, orientation factor (κ^2), is generally constant and assumed to be equal to 2/3. The overall brightness of the FRET pair is another important parameter in choosing a FRET pair because greater brightness

allows more sensitive detection. To increase the brightness, an acceptor with a higher extinction coefficient is also desired.

The considerations described in the previous paragraph apply to the choice of FRET pairs when considering any type of fluorophore, including a FP FRET pair. However, there are additional considerations that are especially or uniquely relevant only to FPs. Generally, FPs are not as robust as synthetic dyes. Some FPs are sensitive to their environment and their fluorescence intensity may depend on external factors such as pH, temperature, and concentration of halide ions. Brightness is a big concern and the tendency of photobleaching is a commonly encountered problem. As a result, great efforts have been invested to engineer new FPs to overcome these problems. It also should be noted that FPs have relatively broad excitation and emission spectra when compared to synthetic dyes, so crosstalk should be taken into consideration as well. Because of these limitations, practical FP FRET pairs are currently limited to: BFP-GFP, CFP-YFP and GFP-RFP, though a few additional FRET pairs have been reported [38-56].

BFP-GFP was the first available FP pair suitable for FRET experiments. The original BFP was obtained by introducing the Tyr66His amino acid substitution into the wild-type avGFP, which results in a blue-shifted of the emission peak from 504nm to 448nm. However, because of BFP's relatively poor extinction coefficient (ε = 13,500 M^{-1}) and quantum yield (ϕ = 0.21) [26], BFP is not an ideal FRET donor. Furthermore, BFP photobleaches very rapidly. Since it was known that the Phe64Leu amino acid substitution improved the brightness of other avGFP variants [27, 57], this same mutation was introduced into BFP [26]. The resulting protein, known as enhanced BFP (EBFP), is five times brighter than its progenitor. The EBFP-EGFP was the only FRET pair option for researchers until the emergence of the CFP-YFP pair. However, EBFP's poor photostability made this FRET pair impractical for many types of applications [4]. Fortunately, several new BFPs, including Azurite and EBFP2 have recently been reported by our group and others [28, 58]. These new BFPs have up to a 500-fold improvement in photostability when compared to EBFP and so these new variants are currently the best FRET donors for EGFP.

The most popular FRET pair reported to date is CFP-YFP. One of the reasons why EBFP undergoes fast photobleaching is that EBFP's excitation spectrum is in the high energy near-UV region. Accordingly, much work has been invested in developing of new FPs with a longer wavelength excitation, so that photobleaching will be less problematic. CFP has an excitation spectrum that is red-shifted about 50 nm compared to BFP, and so it is best excited with visible blue light. For suitable separation of emission peaks, the FRET acceptor of CFP has to be more red-shifted than GFP. The FPs with the most appropriate spectral properties belong to the class of variants known as YFPs. The brightest CFP reported so far is mTFP1 (monomeric teal-colored FP) [59]. The mTFP1 is three-fold brighter than the best BFP ever reported [60]. As a class, YFPs are among the most versatile and brightest FPs yet developed. For example, one of the latest YFP variants, known as YPet (yellow FP for energy transfer), is two times brighter than EGFP, while retaining good photostability [54].

In the past eight years, most attention on FPs has focused on the red spectral regions. The FPs in this class, require longer wavelength green or yellow light for excitation. Thus, they are not only suitable for imaging deeper into tissue (due to greater tissue transparency at longer wavelengths) but also good for the experiments requiring long exposure (due to decreased photodamage). As the furthest red-shifted FPs, RFPs are, in principle, ideally suited for use as FRET acceptors. For example, an RFP acceptor could be paired up with an avGFP donor. In a recent example, protein engineering has been used to create a YFP-RFP FRET pair that is spectrally orthogonal to the mTFP1-YFP FRET pair. These orthogonal FRET pairs can be imaged simultaneously in the same cell [61].

1.2.3 FRET in monitoring protein structure and conformational change

As discussed earlier in this chapter, FRET efficiency is strongly and predictably distance dependent. This was first experimentally demonstrated by Stryer and Haugland in 1967 [37]. They synthesized 12 poly-Pro peptides, varying the number of prolines from 1 to

12. All of the peptides had a naphthyl group on the N-terminus as the FRET donor and a dansyl group on the C-terminus as the acceptor. They measured the FRET efficiency of all the peptides. The measured FRET efficiencies agreed with the theoretical predictions of Förster [34], and showed an inverse sixth power relationship with the distance between the two probes. Numerous researchers have now used FRET as a versatile "spectroscopic ruler" in biological structure elucidation since this seminal work. With the advent of FPs, these chemical-label-free fluorophores paved the way for FRET in becoming multitalented sensors for monitoring protein structures, conformational change, as well as enzyme activities. **Figure 7** serves as a summary of FRET recent applications based on FPs in biological sciences.

1.2.3.1 Intermolecular FRET-based sensors

An intermolecular FRET-based senor, which is well suited for monitoring protein-protein interaction, is illustrated in **Figure 7A**. The two proteins of interest are separately fused with one of the FP FRET pairs. FRET occurs when the two proteins interact because the two FPs are brought into close proximity. Therefore, by monitoring FRET change, the proteins of interest association and dissociation constants can be readily revealed. This strategy has been applied to visualize protein-protein interaction *in vivo*, such as heterotrimeric GTP-binding protein (G-protein) dissociation and reassociation [62], dimerization of receptor protein-tyrosine kinases [63] and homo- and hetero-dimerization of transcription factors [64].

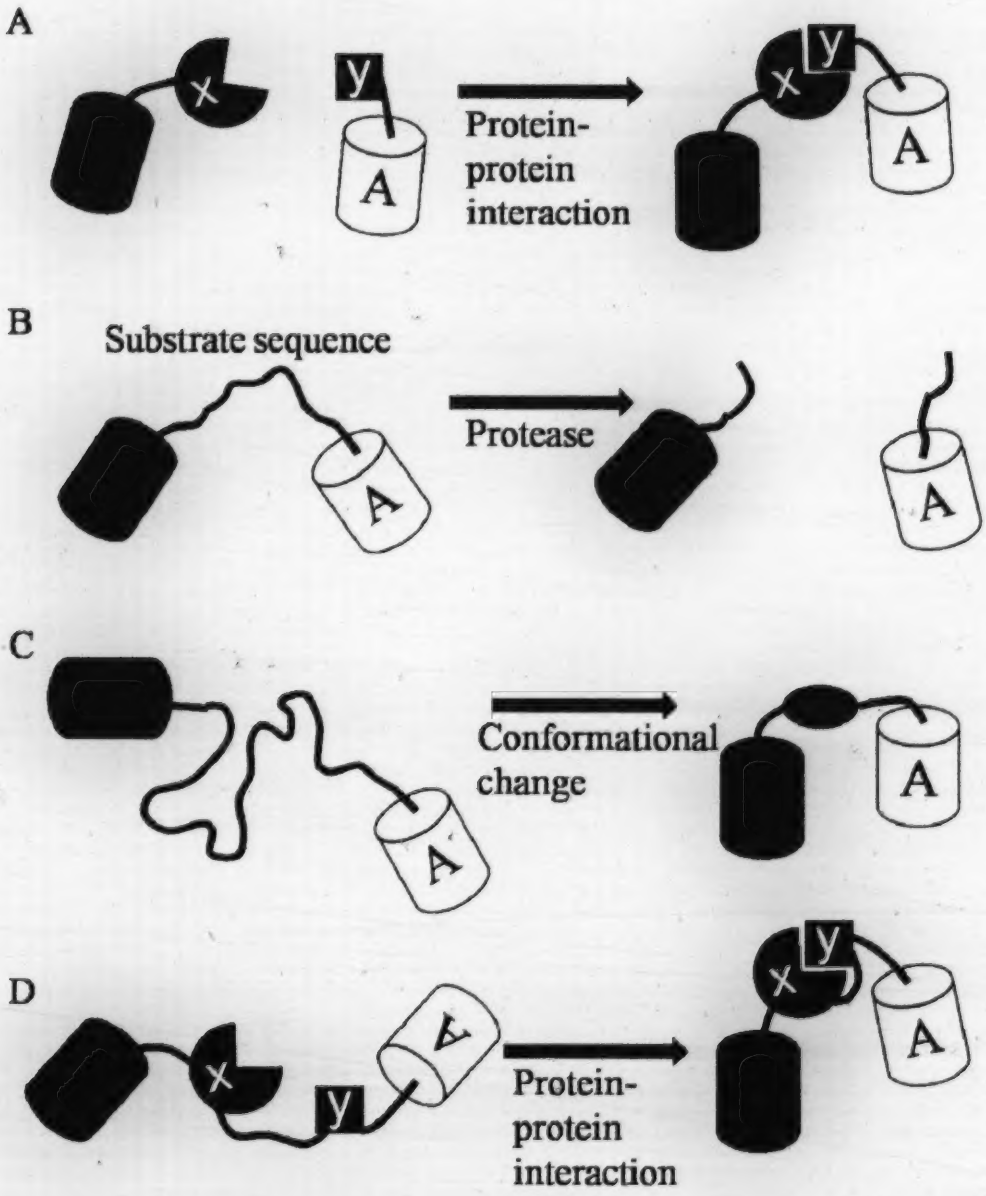


Figure 7 General design of FRET-based FP probes. (**A**) An intermolecular FRET-based FP probe. Two different proteins, protein x and protein y, are separately labeled with the two FPs. One of the FP servers as a FRET donor (D), while the other FP services as a FRET acceptor (A). When protein x and protein y interact, the D and A will be brought to close proximity, which results an increased FRET efficiency. (**B**) An intramolecular FRET-based FP probe for protease activity. The D and A will be apart

after the substrate sequence cleavage, resulting in a decrease of FRET efficiency. (**C**) Intramolecular FRET-based probe for monitoring genetically fused protein conformational change. (**D**) Similar to (**C**), intramolecular FRET can also be used to monitor protein-protein interaction on fusion proteins, such as fusion proteins conformational changed because of post-translational modification or upon small molecule binding.

1.2.3.2 Intramolecular FRET-based sensors

In contrast to intermolecular FRET with FPs, intramolecular FRET with FPs involves the construction of a single gene encoding two differently colored FPs, with one at each end as a "sensing" domain. This type of design lends itself to the creation of FRET-based biosensors that can monitor biochemical processes in live cells. For example, FRET-based protease activity sensors have been created by this strategy [65]. As shown in **Figure 7B**, a substrate sequence, for example, Asp-Glu-Val-Asp (DEVD) is fused between a FRET donor and acceptor. This tetrapeptide sequence is a well known substrate for caspase-3 protease. When the peptide is cleaved, the two FPs dissociate, causes the initially high FRET efficiency to drop to zero. As a result, protease activity can be revealed by monitoring the FRET change either *in vitro* or *in vivo*. Intramolecular FRET-based sensors can also be used to monitor protein conformational change or post-translational modification-dependent protein-protein interactions. As demonstrated in **Figure 7C** and **7D**, single or multiple proteins can be fused between the FRET pair. By monitoring the FRET change, the protein conformational changes and/or protein-protein interactions can be revealed. One of the most successful examples of monitoring protein conformational changes and/or protein-protein interactions is the work of Miyawaki *et al.* [23]. A poly-peptide sequence encoding a calmodulin and the calmodulin binding peptide, M13, was genetically sandwiched between two FPs, which were either BFP-avGFP or CFP-YFP. Calmodulin has high specificity for Ca^{2+} ion. Once calmodulin binds to Ca^{2+} ion, it will be recognized by the M13 peptide, and the overall hybrid protein (calmodulin-M13) will change its conformation from a relatively disordered protein and a

random coil peptide to a highly compacted globular domain. The resulting distance change between the two FPs is elucidated by a FRET change.

So far, most of the proteins that are fused between two FPs are of relatively large size. This is easy to understand: a more noticeable FRET signal change will be obtained when a more dramatic (distance) change occurs. However, through appropriate design of the FP fusion protein construct, this versatile FRET-based strategy can also be applied to monitor the structure of genetically-encoded peptides. This approach will be further discussed in later chapters.

1.3 NOVEL PROTEINS FOR MOLECULAR RECOGNITION

The term "molecular recognition" is usually used in reference to specific shape complementarity and non-covalent molecular interactions between two or more molecules. These non-covalent molecular interactions include: dispersion forces, dipole/induced dipole forces, dipole/dipole forces, hydrogen bonding and electrostatic interactions.

Molecular recognition based on proteins plays a vital role in biological systems as demonstrated by numerous thoroughly investigated and well-understood examples. For example, enzymes, the protein catalysts essential for practically all biological processes, are one of the best examples of molecular recognition. Enzymes are able to recognize their substrates with exquisite specificity. Another good example are immunoglobulin (Ig) proteins (also known as antibodies), which are a key component of the human immune system. Antibodies are vertebrate globulin proteins that serve for identification and neutralization of foreign (non-self) molecules. Antibodies have, in principle, the potential to recognize any provided target. Antibodies with high specificity and affinity can be produced and isolated from the plasma of immunized animals or hybridoma cell lines [66]. The latter approach, which is more preferable in producing large quantities of antibodies, is based on fusing single antibody-forming cells (B-cells) to tumor cells and growing the resulting cells (hybridoma) in culture. For a long time, antibodies were

considered as the best available molecules for recognition of various molecular targets. Generally speaking, this statement is still true, but it is arguable that modern sophisticated protein engineering techniques have produced some non-immunoglobulin proteins for molecular recognition that are competitive or superior to antibodies for specific applications. A variety of novel proteins have been reported that function as specific binding molecules for molecular recognition.

1.3.1 Immunoglobulin or non-immunoglobulin?

It is indisputable that antibodies are the single most successful molecules for molecular recognition in various applications in basic science and medicine: more than 10,000 antibodies are commercially available against a variety of targets (antigens) [67]. In addition, antibodies have emerged in recent years as a distinct new drug class. As of 2006, eighteen antibodies have been approved in the United States for a variety of therapeutic purposes including infectious diseases, chronic inflammatory diseases, transplantation, and cancer treatments. Hundreds of new antibody drug candidates are currently under clinical development [68]. A major reason behind the success of antibodies as human therapeutics is their origin: they are evolved for humoral response (antibodies secretion) against almost any kind of foreign molecules and invading pathogens in higher vertebrates. Accordingly, they are unrivalled in terms of versatility, specificity, and lack of toxicity, compared to drugs based on small molecules or on other protein scaffolds.

The antigen binding regions of antibodies are located at the V_L and V_H domains. The affinity and specificity are mediated by six hypervariable loops known as the "complementarity determining regions" (CDRs) (**Figure 8**). Numerous possible antigen binding sites can be constructed by varying the amino-acid composition in the hypervariable loops, the length of the loops, as well as different V_L and V_H combinatorial association. Prior to about 20 years ago, the vertebrate immune systems were the only sources to generate highly specific molecules for diverse target recognition. Over the last

two decades a number of alternative display and screening technologies, including phage display [69], mRNA display [70]; ribosome display [71] and bacterial surface display [72, 73] have become available.

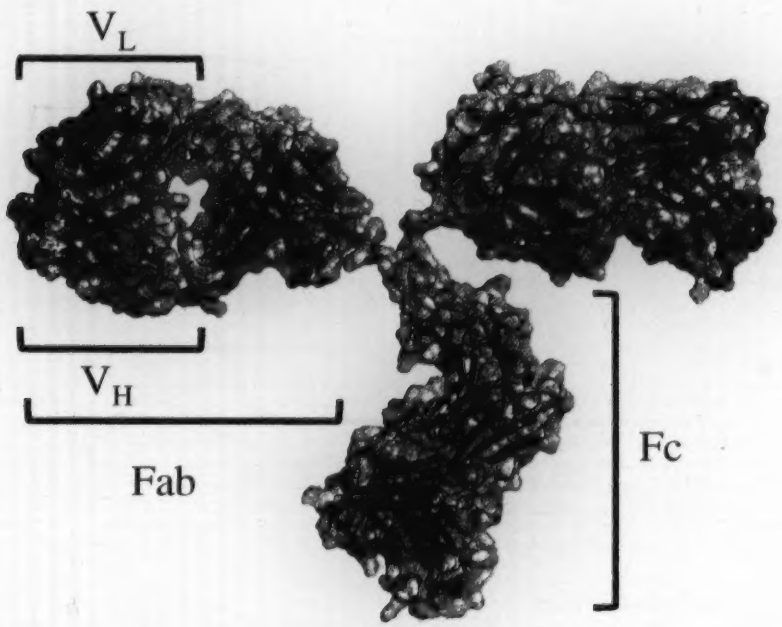


Figure 8 Immunoglobulin G (IgG) structure (PDB:1IGT). The complementarity determining regions (CDRs) are colored in red. Figure prepared with Pymol.

Despite their numerous advantages, antibodies suffer from some disadvantages that limit their usefulness in the increasingly demanding areas of biotechnology and protein therapeutics. Antibodies have an overall "Y" shape structure, and are made of four polypeptide chains; two identical heavy chains and two identical light chains (**Figure 8**). The polypeptide chains are stabilized and connected together by disulfide bonds. These disulfide bonds, which are crucial to antibody structure, make most antibodies unsuitable for intracellular applications in live cells. The cytoplasm of living cells is a highly reducing environment where disulfide bonds are reduced to free thiols due to the presence of thioredoxin reductase and a several millimolar concentration of glutathione. Other features of antibodies, which are often listed as disadvantages, are their relatively

large size (~150 kD) and complicated post-translational glycosylation. Their large structure limits their ability to penetrate tissues, and the large antigen binding sites are inaccessible to some small and buried epitopes. The constant Fc region, which also contributes to the overall large structure, is crucial for generating an appropriate immune response in biopharmaceutical applications, but often leads to undesired interactions in other applications. The complicated glycosylation complicate large scale manufacturing of full-size functional antibodies, as it requires expensive eukaryotic cell culture and delicate optimizations.

Fortunately, with the advent of recombinant DNA techniques and various *in vitro* display strategies, antibodies are no longer the only source of highly specific proteins for molecular recognition. Numerous proteins have now been engineered to have "unnatural" binding ability. Those proteins, which are especially versatile in this regard and can be engineered to bind a variety of different targets, are usually referred as "generic protein scaffolds". The term scaffold came from the elevated platforms that support workers or machines during building construction. But this term "scaffold" adopts another meaning at a molecular level. A generic protein scaffolds is defined as a protein in which residues that confer structural stability are distinct from the residues appropriated as the "unnatural" binding site [74]. In other words, the desired proteins that qualify as generic protein scaffolds should have a significant intrinsic conformational stability so they can tolerate protein engineering purposes, such as amino-acids substitutions, deletions and insertions. Ideally, the desired protein scaffolds should overcome some or all of the drawbacks of antibodies, while not compromising target specificity and affinity.

1.3.2 Novel non-immunoglobulin binding proteins

Non-immunoglobulin binding proteins based on a variety of secondary and tertiary structural motifs will be discussed in this section. I will start by introducing popular protein scaffolds based on fragments or homologs of immunoglobulin domains. I will then move to scaffolds based on α-helix, β-sheet, and combination frameworks. Because I

am primarily interested in applications in the reducing environment of the cytoplasm, the discussion will focus on scaffolds without disulfide bonds.

1.3.2.1 Immunoglobulin-like domains

Fragments and homologs of antibody domains are among the most popular scaffolds under investigation. The majority of research on scaffolds of this type has focused on the antibody fragments that retain the binding specificity and affinity of the whole antibody. The monovalent antigen-binding fragment (Fab) and the single chain variable fragment (scFv) are the most popular engineered formats (**Figure 9A** and **9B**). Fab fragments can be obtained by molecular biology techniques or by digesting intact antibodies with papain to cleave the Fab region of antibodies from the hinge and the Fc region. On the other hand, scFv domains, comprising a V_H and V_L domain that are covalently linked with a polypeptide, can only be generated by molecular biology techniques and recombinant gene expression. There are a range of therapeutic applications that have been reported using Fab and scFv domains, and many of them are in preclinical and clinical trials [75]. However, even though these formats are extremely promising for some applications, poor solubility and decreased binding affinity can cause problems in some applications [76, 77]. Cartilaginous fish (nurse sharks and wobbegong) and camelids (camels and llamas) provide alternative antibody fragments that address some of the problems with antibody fragments [78-81]. Shark Ig-NAR (new antigen receptor) and camelid VhH-Ig (heavy chain antibody VH) are unusual antibodies that contain only one variable domain for antigen binding (**Figure 9C** and **9D**). Compared with Fab and scFv, these variable domains are not only smaller, but also more soluble and show fewer tendencies to form aggregates. However, so-called "humanization" of these new scaffolds is required before they could be employed in therapeutic applications. Humanization refers to grafting the CDRs from the parent domains onto a human antibody framework. Generally, antibody fragments (Fab, scFv, NAR and VhH) provide good sources of relatively inexpensive (*i.e.*, can be produced in microbial hosts) and effective scaffolds for the presentation of constrained polypeptides for molecular recognition, but the overall stability still relies on

disulfide bond formation. As a result, they are not ideal for intracellular applications in live cells.

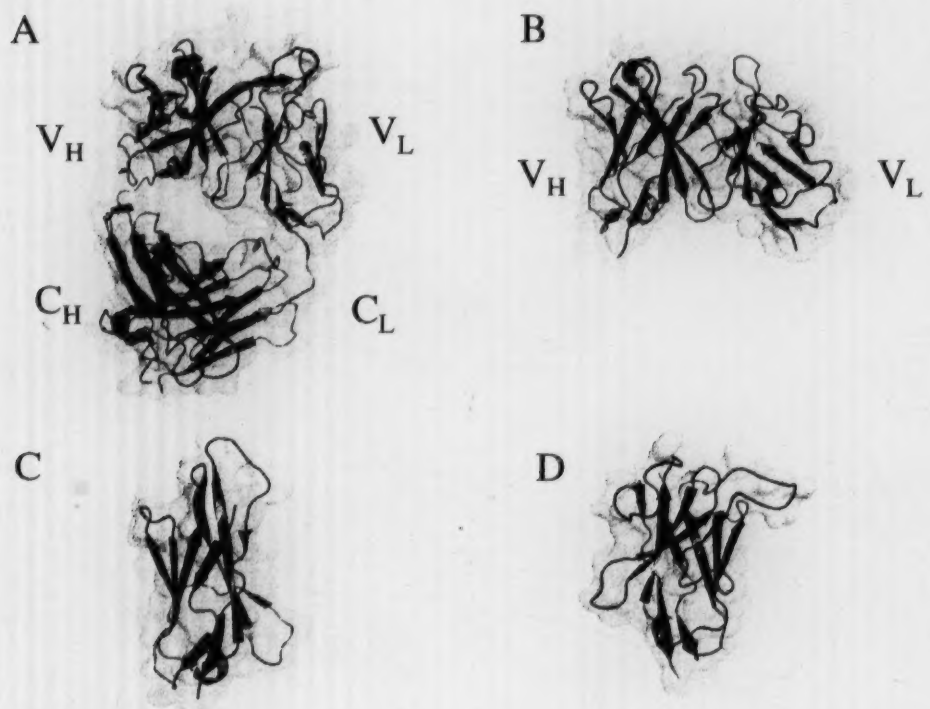


Figure 9 Structures of immunoglobulin-like domains. (**A**) Fab fragment (PDB: 2F58); (**B**) ScFv fragment (PDB: 1N4X); (**C**) Shark Ig-NAR variable domain (PDB: 2COQ); and (**D**) camelid VhH-Ig variable domain (PDB: 1I3V). Unlike typical antibodies, both camelid VhH-Ig and shark Ig-NAR lack a light chain. The CDR regions for molecular recognition in the antibody fragments are colored in red. Graphical representations were prepared by Pymol.

1.3.2.2 Non-immunoglobulin scaffolds with α-helix frameworks

Even though the α-helix is the most abundant protein secondary structure motif in natural proteins, engineered protein scaffolds based on α-helical protein frameworks are

underrepresented in comparison to other classes. Furthermore, the diversity of the α-helical scaffolds that have been reported is generally limited to loops and turns between helices. There are fewer examples in which the randomized residues are located primarily on the helices. There are two possible reasons why scaffolds with randomized residues in α-helices are not particularly common. The first is that a large fraction of the possible substitutions may highly destabilize to the overall structure. The second is that the surface of an α-helix is basically a flat surface that may lack the conformational flexibility for achieving good shape complementarity for high affinity molecular recognition. Nevertheless, a class of high-affinity binding molecules with an α-helix framework, named "affibodies" (**Figure 10A**), are one of the most promising motifs among the non-immunoglobulin binding proteins. Derived from the immunoglobulin-binding domain of staphylococcal protein A [82], affibodies are highly soluble and thermally stable proteins that consist of three α-helices. They possess no disulfide bond and can be effectively produced in bacteria. Engineered as a protein scaffold, affibodies can support randomization at up to 13 positions on two of their three helices. As a result, theoretically huge libraries ($20^{13} = 8\times10^{16}$) could be created for selecting possible binders. Since the first report of using affibodies as protein scaffolds for phage display [83], numerous biotechnological applications of affibodies in detecting and targeting have been reported [84-86]. Affibody-based proteins targeting the HER2 receptor with picomolar affinity have been reported and commercialized (see www.affibody.com) as potential breast cancer diagnostic and therapeutic reagents [87-90]. *E. coli* colicin E7 immunity protein (ImmE7) [91] and cytochrome b_{562} [92] are other examples of protein scaffolds in which the randomized residues are confined mainly in α-helical regions, but these have been the subject of far fewer investigations.

Another approach to engineer protein scaffolds with favorable properties is to exploit protein motifs that can occur in self-associating tandem repeats. In this strategy, the size of the binding interface can be steadily changed by varying the number of repeats. The leucine zipper, which is engineered from a leucine-rich specific-DNA-binding domain, [93] and ankyrin-repeat domain [94] are well studied examples (**Figure 10B** and **10C**). Both have α-helix frameworks, though not all the randomized residues are located in the

helical regions. Lack of disulfide bonds and relatively small sizes make them well suited for intracellular applications. Libraries assembled with 2 to 3 ankyrin-repeat domains and up to 14 repeats of the leucine zipper motifs have been created and selected against a variety of different targets [76, 94, 95].

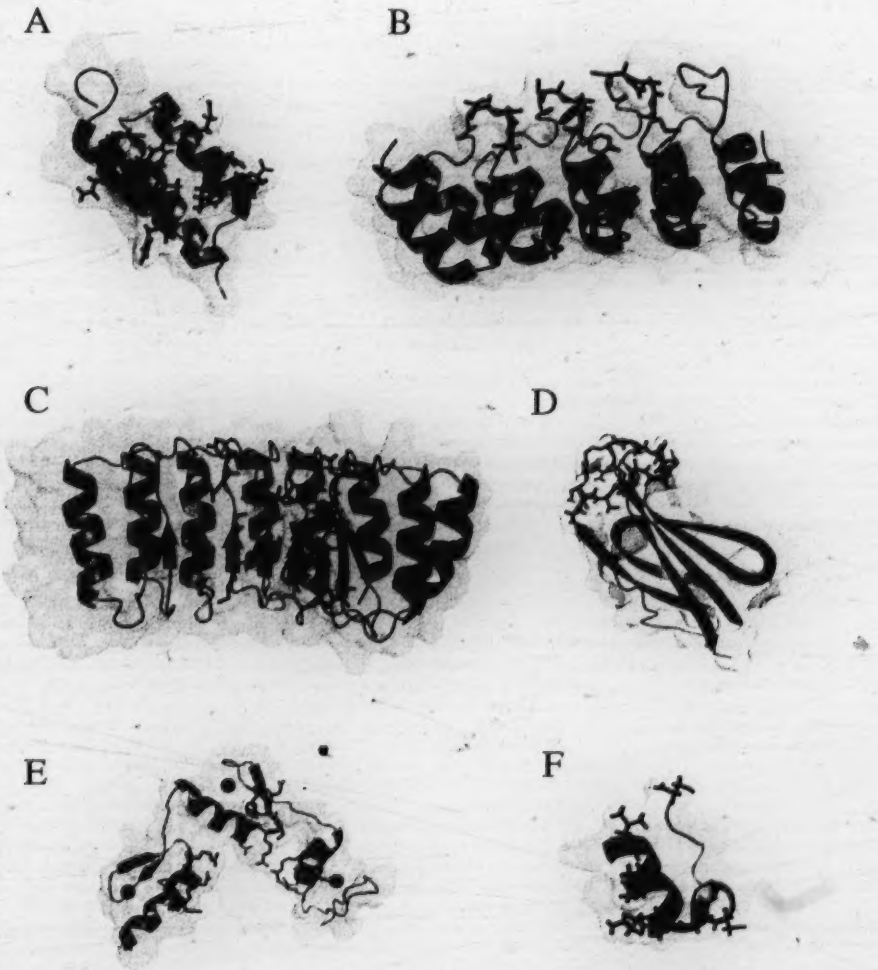


Figure 10 Non-immunoglobulin scaffolds currently used for molecular recognition. The randomized positions are highlighted in red, while the secondary structure of the conserved frameworks of different scaffolds are in green. (**A**) Affibody (PDB:1Q2N); (**B**) ankyrin repeats (PDB: 1SVX); (**C**) leucine zipper repeats (PDB: 1A4Y); (**D**) fibronectin

(PDB: 1FNA); (**E**) zinc fingers (PDB: 1JK1); (**F**) Trp cage motif (PDB: 1L2Y). Graphical representations were prepared by Pymol.

1.3.2.3 Non-immunoglobulin scaffolds with β-sheet frameworks

Antibodies can be considered as the epitome of a protein scaffold with a β-sheet framework. Antibodies can bind to virtually any given target with specificity and affinity that are mediated through the CDRs. These loops are constrained by the β-sheet immunoglobulin domain framework (**Figure 8**). Given the effectiveness of antibodies, considerable effort has been directed toward the development of protein scaffolds that have similar structural properties. However, the disulfide bond(s) is critical for the structural stability in most of the engineered scaffolds based on the β-sheet framework, like the knottins [96] and lipocalins [97]. The scaffolds based on the fibronectin type III domain (^{10}FN3) [98] are one example of a β-sheet scaffold that overcomes some of the limitations of antibodies (**Figure 10D**). Fibronectin is a mostly β-sheet domain that lacks a disulfide bond. It has three surface exposed loops, which are similar in length and orientation to those found in immunoglobulins. The residues in two or three of the loops have been randomized, and selected for binding to various targets by both phage display and mRNA display [98, 99]. Another good example in this category is the "minibody", which was engineered from the immunoglobulin V_H domain [100]. It is interesting that avGFP has also been used as a scaffold for presentation of peptides [101]. avGFP scaffold has a clear and obvious advantage over the classic scaffolds: its intrinsic fluorescence means that it might not function only as a molecular recognition scaffold, but also as a detector provided that a fluorescence change occurs upon binding. Two different approaches have been carried out to display random peptides on avGFP, one is based on rational design [101], while the other approach involved screening of a library of circularly-permuted avGFP variants [102]. It is interesting to note that, in both cases, the regions suitable for randomized peptide insertions are all located in the loop regions rather that the barrel structure of avGFP. Zeytun *et al.* reported taking a more aggressive approach to engineering avGFP to be "fluorobodies" by randomizing 4 of the 6 loops in

avGFP and selecting for molecular recognition properties using phage display [103]. However, that work was later retracted [104].

1.3.2.4 Non-immunoglobulin scaffolds with combined structures

Proteins within this group of scaffolds comprise of a mixture of α-helices, β-sheets, and loops. The majority of scaffolds fall into this category, ranging from small peptides such as the Trp cage motif [105] and zinc finger motif [106] to relatively large proteins like serum transferrin [107] and β-lactamase [108]. One of the most important protein scaffolds in this category, the zinc finger proteins, have emerged as programmable scaffolds for specific nucleotide sequence recognition [109]. First identified in 1985 [110, 111], zinc finger proteins soon proved to be an ideal model system for the study of sequence-specific DNA binding. Zinc fingers are short peptides of ~26 amino acids. Their structure is composed of two antiparallel β strands followed by an α helix (**Figure 10E**). The zinc finger tertiary structure arises from the common Cys_2His_2 motif, which binds to a zinc ion. The sequence-specific binding of DNA is mediated by the amino acids on the α-helix. Each zinc finger can specifically recognize and bind to three contiguous DNA base pairs in the major groove of double stranded DNA (dsDNA). Ideally, if zinc fingers that recognize all possible sequences of three DNA base pairs were discovered in nature, they would be great building blocks for creating proteins that can recognize any given DNA sequence. Unfortunately, nature does not provide us with all the necessary building blocks. Therefore, different research groups carried out studies on engineering new zinc fingers that have new nucleic acid sequence specificities [112-115]. Today, engineered zinc fingers can recognize all the continuous or separated DNA triplets, and the amino acids of these zinc fingers can be easily obtained using software programs available online [116].

It is well understood that a rigid scaffold can provide higher affinity binding because of decreased entropic cost upon binding to the targets. The Trp cage motif is one of the few mini-protein scaffolds with a rigid structure that is achieved without the assistance of

disulfide bond(s) (**Figure 10F**) [117]. Instead, the structural stability of the Trp cage arises from hydrophobic interactions with a tryptophan side chain buried in the centre of the structure. In an attempt to use the Trp cage as a protein scaffold, seven of the solvent-exposed positions were randomized and screened for streptavidin and human bronchial epithelial cells binding by phage display [105]. The promising results showed that the Trp cage has great potential to be a mini-protein scaffold for target-specific molecular recognition.

β-lactamase is another good example of a relatively large scaffold in this category. β-lactamase is an enzyme that hydrolyzes β-lactam antibiotics such as penicillin. Randomized sequences were introduced around the active site of β-lactamase, and the resulting protein libraries were screened for binding horse spleen ferritin [108]. The enzyme activity changed upon binding thus enzyme activity could be used as an indication of target binding. But unlike the avGFP scaffold, the clear drawback of this approach is that β-lactamase cannot be applied in cytoplasm, because of its inherent and essential disulfide bonds.

1.3.3 Phage display libraries and their screening

As should be evident from the previous sections, identifying a potential protein scaffold is only the first step in creating novel binding proteins with useful molecular recognition properties. It is also necessary to create genetic libraries of variants and then to identify those rare variants that have the desired binding property. Accordingly, the single most important breakthrough in engineering novel binding proteins was the advent of display and selection technologies. Engineering new properties into an existing protein can be achieved by generating a pool of mutants and screening for the desired characteristics. By recombinant DNA techniques, trillions of mutants can be obtained readily, however, it is impossible to characterize all individual mutants one at a time even with current and sophisticated high-throughput methodologies. Consequently, technologies that enable the rapid selection of only the desired candidates from the pool offer the best solution.

A selection technique relies on the establishment of a linkage between the phenotype and genotype. These linkages are a physical coupling between the protein and the encoding gene, so the protein sequence of the variant with the desired property could be readily determined by a DNA sequencing reaction. Both *in vivo* and *in vitro* selection technologies are available. The selection can be performed in live cells (e.g., yeast surface display and bacterial surface display [72, 73]) or viruses (e.g., filamentous phage display and [69] T7 phage display (Novagen)). Alternatively, they can be done on DNA-protein or mRNA-protein complexes made by *in vitro* translation in a cell-free system (e.g., ribosome display [71], mRNA display [70] and DNA display [118, 119]). Among all of the different selection strategies, phage display is by far the most widely used method.

Phage display, as the name implies, involves the display of single or multiple copies of the desired peptide or protein on the surface of phage particles. The bacteriophage used for selection is usually from the Ff group of filamentous phage (f1, fd and M13), though techniques with T7 (Novagen) and λ phage (Stratagene) are available. The Ff group represents two very important properties: one is the phage filamentous appearance; the other is the infection dependence on the *E. coli* host F pilus. Ff phage are non-lytic viruses that selectively infect host bacteria with the F pilus. The F pilus is a protein tube that extends from the bacterial cytoplasmic membrane into the medium. Only bacteria with the F plasmid DNA have the F pilus. Thus, Ff filamentous phage are specific for *E. coli* carrying an F plasmid.

In the core of the Ff phage particle is the single-stranded circular genome, which is encapsulated in a protein cylinder (**Figure 11**). In the typical case, the 930 nm long protein cylinder consists approximately of 2700 copies of the 50-amino acid α-helical *gene VIII* protein (pVIII). The pVIII molecules are packed together very tightly such that only 3 of the N-terminal amino acids are exposed to the surface and are sensitive to digestion with a protease [120]. On one end of the phage particle are 3-5 copies of *gene VII* protein (pVII) and *gene IX* protein (pIX). pVII and pIX are less well studied than other protein components of the phage particle, and neither their structures nor their exact

arrangement is known. However, they are critical for phage formation because it is believed they are involved in initiating phage particle assembly, and no phage particles will form if either protein is absent. On the other end of the phage particles are three to five copies of *gene III* protein (pIII) and *gene VI* protein (pVI). pVI is less well studied, but has been found that to be buried within the phage particle [121]. pIII is made up of three domains (lower left in **Figure 11**), which are separated by glycine-rich linker regions. The first two domains, N1 and N2, are critical for phage infection. The N1 domain is responsible for phage DNA translocation while the N2 domain is for F pilus recognition. The third domain, CT domain, together with pVI, forms one end of the phage particle by interacting with pVIIIs.

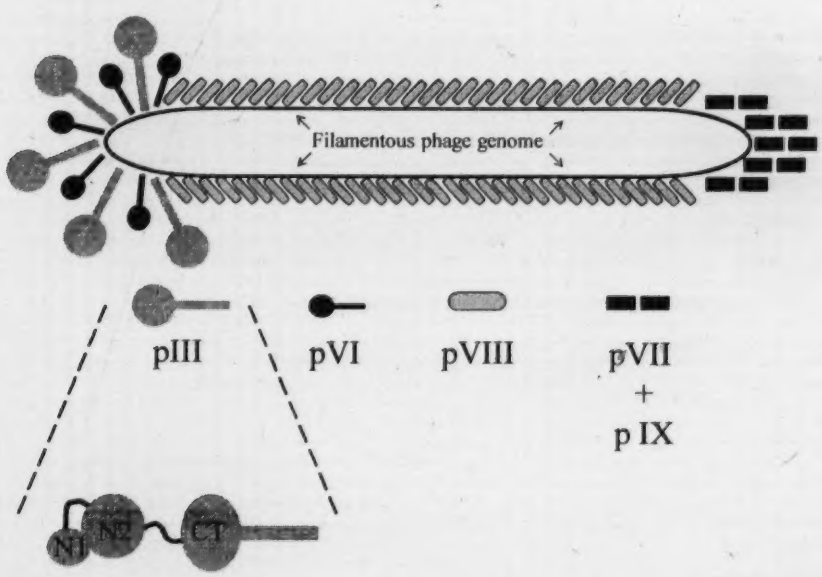


Figure 11 Schematic representation of the filamentous baceteriophage particle. At the lower left is a more detailed representation of the *gene III* protein (pIII).

The phage life cycle is a multistep process, which includes infection; DNA replication; protein synthesis; and phage particle assembly. The first step of infection involves the binding of phage pIII N2 domain to the F pilus of the *E. coli* [122, 123]. The F pilus then

retracts and brings the phage close to the bacterial periplasm. This allows the N1 domain to interact with the bacterial membrane protein, TolA, to initiate the infection, but subsequent steps are unclear [124]. Once the viral plus-stranded (+) ssDNA is translocated into the cytoplasm, phage DNA replication and protein synthesis start (**Figure 12**). The complementary minus-stranded (-) ssDNA is synthesized, supercoiled, forming dsDNA with (+) ssDNA. This dsDNA is known as replicative form DNA. The (-) ssDNA is the template for phage protein synthesis. The dsDNA synthesis continues until the expression of *gene V* protein (pV) reaches a critical concentration. The pV dimers interact with (+) ssDNA, thus preventing dsDNA formation [125]. Those (+) ssDNA will later be encapsulated into the phage particles as viral genomes. Phage coat proteins (pIII, pVI, pVIII, pVII and pIX) together with three assembly proteins *gene I* protein (pI), *gene IV* protein (pIV) and *gene X* protein (pX) are synthesized and translocated mainly to the cytoplasmic membrane (**Figure 12**) [124]. One exception is pIV, 12-14 copies of which form a cylindrical channel structure or pore located in the outer membrane, and functions for phage particle secretion [126]. The other two assembly proteins, pI and pX, are believed to form a channel in the cytoplasmic membrane [127], although such a complex have not yet been isolated. The phage assembly process is initiated by the interaction between the (+) ssDNA and the pVII, pIX, as well as pVIII [128]. But the protein-DNA interaction arrangements are not known yet [124]. After initiation, the pre-phage-particles undergo an elongation process by replacing pV, which is attached to the ssDNA, with the major coat protein pVIII. The pre-phage-particle continually elongates as it is secretes from the channels formed by pIV, pI and pX. When the end of ssDNA is reached, pIII and pVI are added to the end of the phage particle to terminate the assembly process.

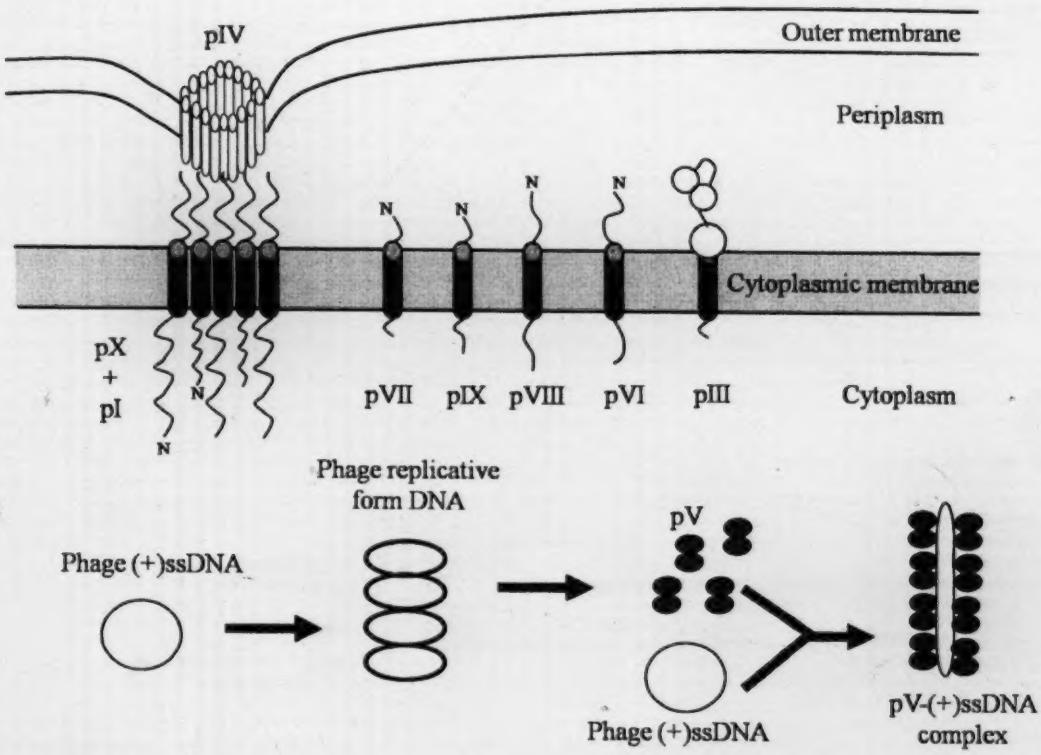


Figure 12 A schematic representation of the phage DNA replication, protein synthesis and protein location in a Ff phage cycle (modified from Carlos *et al.* [125]).

As with any efficient display technique, the phage display strategy involves linking the genotype and phenotype together. This is by fusing the nucleotide sequences encoding the desired display peptide or protein library with the gene encoding a phage coat protein. The fusion gene is inserted into a phage genome or a phagemid plasmid. This fusion ensures that the DNA sequence is contained in the phage particles while the encoding peptide or protein library is displayed on the phage particles surface. All of the five capsid proteins have been used for displaying either peptides or proteins; however, the dominant virion proteins that have been used are pIII and pVIII.

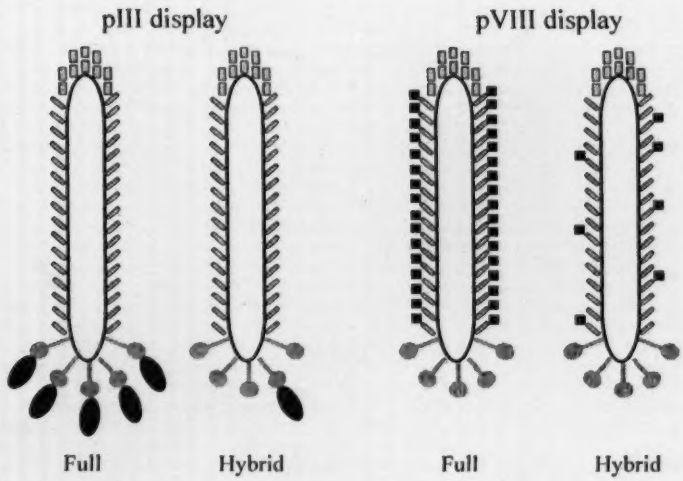


Figure 13 Schematic representation of foreign protein displayed on the surface of filamentous phage. The dark symbols represent the foreign peptides or proteins as fusions to the pIII and pVIII.

Generally, phage display can be divided into two different groups: full display or hybrid. Both types have been used in the display of proteins on pIII or pVIII (**Figure 13**). In full display, the peptide/protein sequence is genetically inserted into a natural Ff phage genome. As a result, every coat protein, pIII or pVIII in this case, is a chimera fused with the desired peptide or protein. Thus, 3-5 copies of the desired peptide or protein will be on the surface of a phage particle in the pIII display, while more than 2000 copies will be presented when pVIII display is chosen! The major drawback for this full display approach is the size limitation of the display peptides or proteins. It has been reported that only peptides less than 8 amino acids long can be displayed on every copy of pVIII [129, 130]. One reason may be that the large inserted peptide or protein adversely affect the assembly of the phage particles [124]. Size problems are also associated with the full pIII display. Larger proteins can significantly lowered phage infectivity, and in some cases make the phage noninfective.

An alternative approach is the hybrid system which uses so-called "phagemid" vectors. The phagemid approach is now the strategy most commonly reported in the literature for protein display. Phagemid vectors contain the fusion protein gene and no other phage genes. Upon transformation of *E. coli* bacteria with the phagemid, the fusion gene will be expressed and the fusion proteins will be abundantly present in the *E. coli* cells. The phagemid vector also has an Ff origin to allow phagemid ssDNA production. The bacteria then are infected by "helper" phage, which carry the full phage genome. Once inside the bacterium, the genes on the helper phage genome are expressed to produce all the necessary "wild-type" proteins for new phage particles assembly. However, the abundant chimeric coat proteins, resulting from expression of the phagemid fusion gene, compete with wild-type coat proteins for phage assembly. Furthermore, because the helper phage genome has a compromised origin that leads to inefficient packing, the majority of ssDNA inside of the phage particles are the phagemid vectors. For most of the cases, only one copy (monovalent) or less of the desired peptide or protein fusion with pIII is incorporated into the phage particle. Likewise, only a relative few copies are incorporated in pVIII display (polyvalent). The main advantage of the phagemid hybrid system is overcoming the limitation of display protein size. Numerous large proteins, such as antibody fragments and enzymes, have been displayed efficiently on phage surface using the hybrid approaches. In addition, monovalent peptide or protein display using the pIII hybrid system allows selection based on monovalent affinity. It has been noted that polyvalent selection prevents the highest-affinity binders to be discovered [131]. This is because multivalent screening (full display and pVIII hybrid) is based on avidity rather than affinity. As a result, pIII hybrid display is preferred for identification of the highest affinity variants.

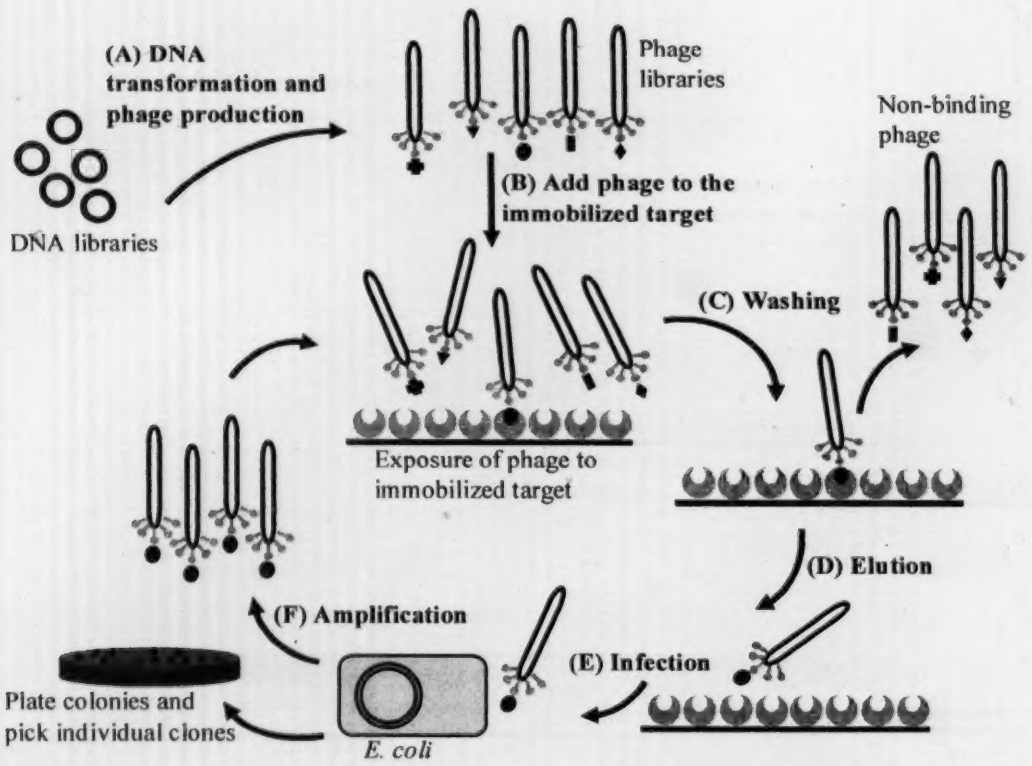


Figure 14 Screening a phage displayed library. (**A**) DNA libraries encoding the displayed peptide or protein are used to produce phage libraries in *E. coli*. The phage displaying the desired peptide(s) or protein(s) are harvested and (**B**) exposed to the immobilized target. Followed by washing steps (*e.g.* using phosphate buffered saline (PBS)) (**C**), non-binding phage are washed off. (**D**) Bound phage are eluted by either disrupting the peptide(s)/protein(s)-target interaction (*e.g.* low pH buffer) or adding known competitive molecules to the immobilized target. (**E**) The eluted phage are used to infect *E. coli* to produce offspring, therefore, the phage are amplified and ready for subsequent selection (**F**). The infected bacteria can also be plated as colonies on culture plates. Individual clones will be picked, and the plasmid DNA will be purified and sequenced.

Phage library screening procedures are relatively simple and amenable to parallel screens against a variety of targets (**Figure 14**). In addition, it is fairly rapid, with one cycle of screening taking about a day to complete. As a rough guide, millions of variants could be

screened within a week. Since first introduced in 1985 [132], phage display has been one of the most important combinatorial biological methods. Numerous reports of antibody engineering have been based on the phage display library screening. It also has had a tremendous impact on novel binding protein discovery: more than 70% of protein scaffolds were screened for molecular recognition based on phage display [76]. Phage display has also been developed as a valuable methodology in enzyme evolution [133].

1.4 SUMMARY

This chapter provides a brief introduction of important concepts and methodologies throughout this thesis. A more detailed discussion of FRET in monitoring genetically fused peptide structure will be served in later chapters. The phage display techniques will be applied in screening streptavidin binders, which will be discussed in **Chapter 5**.

CHAPTER 2:

ASSESSING THE STRUCTURAL STABILITY OF DESIGNED β-HAIRPIN PEPTIDES IN THE CYTOPLASM OF LIVE CELLS

The work of this chapter was published in as a research paper "Assessing the structural stability of designed beta-hairpin peptides in the cytoplasm of live cell", *Chembiochem*, 2006, **7**: 1147-1150.

2.1 INTRODUCTION

Ribosomally-synthesized peptides and proteins that are capable of specific molecular recognition in the cytoplasm or nucleus of the living host cell are valuable research reagents that provide an effective means to inhibit or modulate specific pathways [134, 135], or localize specific components [136, 137]. The vast majority of published research directed towards the development of such reagents has focused on the use of recombinant antibody fragments [138], so called intrabodies, that can be expressed in living cells and retain the binding specificity of the intact antibody from which they were derived. The two major limitation of intrabodies are their size and their notoriously poorly folding efficiency in the reducing environment of the cytoplasm, due to an inability to form the intra-domain disulfide bond, which is critical for their stability [139]. One approach towards addressing this limitation has been to engineer alternative protein structures that lack disulfide bonds and engineer them to have "unnatural" molecular recognition functions [74, 140].

The proteins best suited for conversion into molecular-recognition domains have been designated generic protein scaffolds. The desirable features of a generic protein scaffold intended for intracellular applications include a relatively small size (to reduce non-specific binding and allow access to epitopes that are structurally occluded from larger domains), and the absence of a disulfide bond (discussed above). To date, the smallest validated generic protein scaffold that meets these criteria are zinc-fingers at just 26 amino acids in length [106]. Despite their small size, these mini-proteins meet the basic criteria of a generic protein scaffold because the part of the structure that confers structural stability (the Cys_2His_2 coordination of a zinc ion) is completely independent of the part the structure used for the molecular recognition function (5 contiguous residues

on the face of an α-helix). Zinc-fingers are known to be expressed and fold intracellularly so variants derived from such libraries should be suitable for molecular recognition in the cytoplasm or nucleus [109]. An alternative approach for display of random loops in a "stem-loop" peptide structure [141] lacking disulfide bonds and just 31 amino acids in length has recently been described [142]. Following the successful panning of the library to find high affinity binders of the Texas red fluorescent dye, the authors went on to demonstrate that the selected "stem-loop" peptide sequences retained their binding function when intracellularly expressed. In an effort to add a new minimal motif to the growing repertoire of validated non-immunoglobulin binding proteins, we have undertaken the development and validation of a generic protein scaffold based on a single β-hairpin that can fold efficiently in the cytoplasm.

β-hairpin peptides of the "tryptophan zipper" (trpzip) type are promising candidates for a minimal generic protein scaffold. Notably, trpzip peptides are only 16 to 20 amino acids in length, monomeric, and highly soluble. Trpzip peptides owe their origins from the B1 domain of the IgG binding protein (GB1p) in *Streptococcus* [143]. The GB1p consists of one α-helix and two β-hairpins. Despite the overall small size and absence of any disulfide bond, GB1p has considerable resistance to heat and urea denaturation [144]. This remarkable stability results from its secondary structure, particularly from residues 41-56 which forms the second β-hairpin structure. The residue sequence, GEWTYDDATKTFTVTE was identified, synthesized and characterized [145]. It was found that, even without the context of the remaining residues in GB1p, the peptide still adopts a β-hairpin conformation in aqueous solution. The peptide structure is attributed to the hydrophobic cluster formed by the valine side chain and the aromatic rings from the Trp, Tyr and Phe residues (underlined). A significant stability improvement was witnessed by replacing all four side chains with indole rings (Trp): the resulting peptide, named **trpzip4**, was reported to have a melting temperature greater than 70 °C [146]. NMR structures revealed that the four Trp residues form two interdigitating cross-strand Trp-Trp pairs on one face (the Trp-face) of the hairpin. It is believed the strong $\pi - \pi$ interactions from these two pairs of Trp residues gives rise to the remarkable thermal stability of **trpzip4** peptide. Further optimizations were also carried out in the loop region

to further improve trpzip stability [146]. The trpzip peptide we studied is **HP5W4** (KKWTWNPATGKWTWQE) (**Figure 15**), with a reported melting temperature of 85 °C [147].

We hypothesized that residues with side chains directed towards the non-Trp-face and in the loop regions are less important for trpzip β-hairpin structural stability. If so, it may be possible to select structured β-hairpins that are preorganized for target binding from phage displayed libraries [69] in which multiple non-Trp-face residues have been randomized. Related β-hairpin sequences that contain Trp-Lys cross strand pairs on the Trp-face have recently been shown to selectively bind ATP [148], flavin [149], and ssDNA [150]. Alternatively, considering the great stability of trpzips in forming β-hairpin structures, they might be particularly amenable to conversion into "stem-loop" structures by replacement of their turn sequences with longer randomized residues.

If desired, the binding affinity could be enhanced by adopting the repeated-protein-motif strategy, like the ankyrin repeat domain [94] and the leucine zipper [93] discussed in **Chapter 1**. Polypeptides composed of multiple linked β-hairpins could serve as the basis for an unnatural class of multivalent binding proteins that would be suitable for use in the reducing environment of the cytoplasm. As such, these proteins could serve as an intracellular complement to "avimers"; a recently-reported class of high affinity binding modules that is based on multiple linked domains of ~35 amino acids residues each [151]. These multivalent proteins are capable of sub-nanomolar binding affinities against a variety of protein targets though they are stabilized by disulfide bonds and therefore are unsuitable for use in the cytoplasm.

A desired generic protein scaffold should be able to tolerate multiple amino acid substitutions without significant structural destabilization. Therefore, the first question we attempted to answer was whether the trpzip (**HP5W4**) is indeed a good scaffold in withstanding amino-acid substitutions. The conventional approach for development of β-hairpin peptides generally involves chemical synthesis of a number of rationally and/or systematically-modified peptide variants that are individually characterized by circular dichroism (CD) and/or NMR spectroscopy [146, 152-155]. A combinatorial approach

based on the CD spectroscopy-guided deconvolution of peptide mixtures has also proven an effective means of developing peptide sequence that fold into a β-hairpin peptides [156, 157]. A recently reported approach for characterization of β-hairpin peptide structure is to synthetically tether both a fluorophore and a quencher moiety onto the peptide and to use the rate of end-to-end collisions to study dynamic flexibility [158].

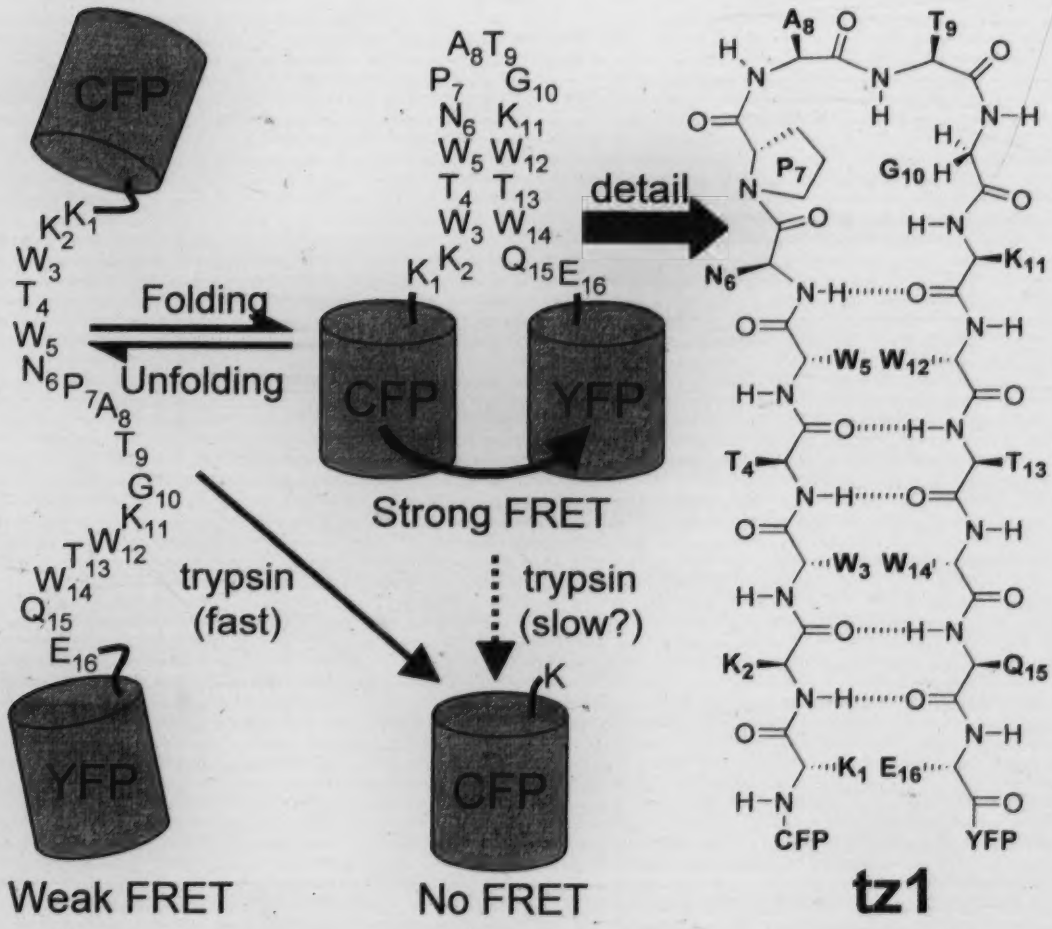


Figure 15 Schematic illustration of the FRET efficiency and rate of trypsin digestion. Both of them depend on the amount of β-hairpin structure in the peptide portion of the Nterm-CFP-peptide-YFP-Cterm proteins.

Although these previously-reported approaches have proven highly successful at developing highly stable β-hairpin peptides, they cannot be readily adapted to the study of β-hairpin peptides in the cytoplasm of living cells. We have therefore developed a versatile method for rapidly evaluating the β-hairpin structure of recombinant peptides both *in vitro* and *in vivo*. Our strategy is based on the presumption that attaching a cyan fluorescent protein (CFP) and a yellow fluorescent protein (YFP) to the ends of an intervening peptide sequence would provide a convenient fluorescence resonance energy transfer (FRET) based probe of the end-to-end distance (**Figure 15**). This approach is conceptually analogous to the "molecular beacon" approach for *in vitro* detection of specific nucleic acids [159]. It was anticipated that the FRET efficiency would be higher for peptides that exist in a predominantly β-hairpin conformation than for unstructured peptides of the same length. A major advantage of using a genetically encoded FRET pair is that new peptide sequences can be rapidly created and screened through the combination of molecular biology and fluorescence imaging of bacterial colonies. A logical extension of such a method would be the high-throughput screening of large libraries of peptide sequences for β-hairpins that are highly structured in live cells.

2.2 MATERIAL AND METHODS

2.2.1 General procedures and materials

The gene encoding the CFP was constructed by introducing the Cerulean [160] mutations encoding the following amino acid substitutions: Ser72Ala, Tyr145Ala and His148Asp into the gene encoding for ECFP (avGFP with mutations encoding Lys26Arg, Phe64Leu, Ser65Thr, Tyr66Trp, Asn146Ile, Met153Thr, Val163Ala, Asn164His, and His231Leu) [23] in the pBAD/His B vector (Invitrogen Corp., Carlsbad, CA) by a whole-plasmid amplification-based method using mismatched primers. This specific version of CFP has not been previously described in the literature but is at least as brightly fluorescent as Cerulean (Hui-wang Ai and Robert E. Campbell, unpublished results). The gene encoding the YFP is the Ala206Lys mutant [17] of Citrine [161]. All synthetic DNA

oligonucleotides were purchased from Sigma-Genosys Canada (Oakville, ON). PCR products and products of restriction digest were routinely purified using the QIAquick PCR purification kit according to the manufacturer's protocols (Qiagen, Mississauga, Ontario). Restriction enzymes were purchased from either Invitrogen or New England Biolabs (Ipswich, MA). The identity of all cDNA constructs was confirmed by dye terminator cycle sequencing using the DYEnamic ET kit (Amersham Biosciences, Uppsala, Sweden). All sequencing reactions were analyzed at the University of Alberta Molecular Biology Service Unit. Unless otherwise noted, protein samples for all biochemical characterization experiments including FRET measurements, proteolysis, and CD spectroscopy were in 50 mM Tris-HCl, pH 7.5. To calculate CFP to YFP interchromophore distances (r) from FRET efficiencies (E), the following formula [162] was used with $R_0 = 5$ nm: $r = \left(\frac{R_O^6}{E} - R_O^6 \right)^{\frac{1}{6}}$ [55].

2.2.2 Construction of CFP-peptide-YFP expression vectors

Construction of CFP-peptide-YFP expression vectors: the vector used for cloning and expression of all peptides flanked by CFP and YFP is referred to as pZC1 and was constructed as follows. The gene encoding YFP was PCR amplified with *Pfu* DNA polymerase (Fermentas Canada Inc, Burlington, ON) and a 5' primer that appended *Xho*I/*Kpn*I/*Pst*I restriction sites to the codon encoding Glu5. The 3' primer appended an *EcoR*I site after the stop codon. The PCR product was digested with *Xho*I/*EcoR*I and ligated with T4 DNA ligase (Invitrogen) into pBAD/His B (Invitrogen) that had been digested with the same two restriction enzymes. The ligated product was used to transform *E. coli* DH10B (Invitrogen) by electroporation. Appropriate dilutions of transformed bacteria were plated on LB/agar containing ampicillin (0.1 mg/ml) and arabinose (0.2%) and after overnight incubation at 37 °C, single colonies were picked and used to inoculate 2 mL of LB media containing ampicillin (0.1 mg/mL). Cultures were grown with shaking overnight at 37 °C before plasmids were isolated using the QIAprep

spin miniprep kit according to the manufacturer's protocols (Qiagen, Mississauga, Ontario). The gene encoding CFP was PCR amplified with a 5' primer that appended an *Xho*I restriction site before the start codon. The 3' primer appended the *Spe*I/*Sac*I/*Kpn*I restriction sites immediately after the codon encoding Thr230. The PCR product was digested with *Xho*I/*Kpn*I and ligated into the similarly digested pBAD/His B vector containing the gene for YFP. Bacterial transformation and plasmid isolation was performed as described above. The resulting vector (pZC1) has a convenient set of restriction sites (*Spe*I/*Sac*I/*Kpn*I/*Pst*I) flanked by a 5' gene for CFP and a 3' gene for YFP with a minimal intervening linker sequence. dsDNA encoding the target peptides and with appropriate sticky ends for ligation into the *Sac*I/*Kpn*I sites of pZC1 were created by slowly cooling pairs of complementary single stranded oligonucleotides from 95 °C to room temperature. The synthetic annealed dsDNA sequences were ligated into pZC1 that had been digested with *Sac*I and *Kpn*I. Bacteria transformation and plasmid isolation was performed as described above. The sequence of the insert was confirmed by DNA sequencing with a forward primer (5'-CCCTCGTGACCACCCTGACCTGG-3') that anneals to the chromophore region of the gene for CFP but not YFP. The resulting plasmids encode proteins of the general structure N_{term}-His6-EK-CFP(1-230)-TSGAQ-peptide-GTSAE-YFP(5-238)-C_{term}, where "His6" represents six consecutive histidine residues that facilitate metal affinity purification [163], "EK" is the recognition sequence for the protease enterokinase, "TSGAQ" is the sequence (actagtggagctcag) contains both a *Spe*I and a *Sac*I restriction site, "GTSAE" is the sequence (ggtacctctgcagag) contains both a *Kpn*I and a *Pst*I restriction site and "peptide" is a 16mer from the **tz** series.

To construct the CFP-**tz1**-CFP and YFP-**tz1**-YFP expression vectors, pZC1 containing the gene for full-length **tz1** was digested with either *Xho*I/*Sac*I or *Kpn*I/*EcoR*I to excise the gene for CFP or YFP, respectively. In place of the excised fluorescent protein gene was inserted the PCR amplified and appropriately digested gene for YFP or CFP, respectively.

2.2.3 Construction of CFP-tz1-peptide expression vector

To construct the vector for expression of a CFP-fused protein precursor of the **tz1**-peptide, pZC1 containing the gene for the full-length **tz1** protein was digested with *Kpn*I/*EcoR*I to remove the YFP gene. The digested vector was ligated with a synthetic dsDNA that had appropriate sticky ends and a stop codon immediately after the *Kpn*I site. The resulting vector was subject to whole plasmid PCR amplification with a 5' (sense) primer that annealed immediately 5' of the **tz1** sequence and an appended *Sac*I restriction site. The 3' (antisense) primer annealed to the 3' of the gene encoding CFP and appended a sequence encoding an AcTEV protease recognition sequence (ENLYFQG) followed by a *Sac*I restriction site. The resulting PCR product was digested with *Sac*I and circularized by self-ligation. The resulting vector encodes proteins with the general structure Nterm-His6-EK-CFP(1-230)-AcTEV-AQ-peptide-GT-Cterm, where "AcTEV" is the protease recognition sequence, and "peptide" is the peptide portion of protein **tz1**.

2.2.4 Protein and peptide expression and purification

The pBAD/His B vector used for all cloning and expression allows proteins to be expressed at high levels under control of the arabinose operon. A typical protein purification procedure started with the inoculation of 1 L LB media containing ampicillin (0.1 mg/mL) and arabinose (0.2%) with a single colony of *E. coli* DH10B expressing the protein of interest. Cultures were grown overnight at 37 °C with shaking at 225 rpm. Cultures were cooled to 4 °C on ice and harvested by centrifugation (10 min, 4000 rcf). The cell pellet was resuspended in PBS buffer (130 mM NaCl, 2.7 mM KCl, 10 mM phosphate, pH 7.6) and the cells lysed by a single passage through a French Press (Thermo Electron, Waltham, MA). Insoluble cell debris was pelleted by centrifugation at 4 °C (10 min, 15,000 rcf) and 1 ml of Ni-NTA resin (Qiagen) was added to the supernatant. Following one hour of gentle mixing at 4 °C, the supernatant was loaded onto a 6 ml polypropylene column, washed, and gravity eluted in 100 mM imidazole (pH 7.5) according to the manufacturer's protocol. FRET constructs containing both CFP and

YFP were further purified on an AKTAdesign chromatography system equipped with a Hiload 16/60 Superdex 75 prepgrade gel filtration column (Amersham Biosciences, Buckinghamshire, England) that was equilibrated with 50 mM Tris-HCl pH 7.5. Isolated proteins were concentrated with a Centricon centrifugal filter YM-30 (Millipore, Billerica, MA) and stored at 4 °C.

For the CFP-**tz1**-peptide protein, the Ni-NTA purified protein (7.5 mg, 0.25 μmol) was treated with 50 units of AcTEV protease (Invitrogen) and incubated at 30 °C for 4 days. Following complete digestion, the released peptide (sequence Nterm-GAQKKWTWNPATGKWTWQEGT-Cterm, designated **tz1**-peptide), was purified by reversed-phase chromatography on an AKTAdesign chromatography system equipped with a Source15RPC ST 4.6/100 column (Amersham Biosciences). A linear gradient of increasing acetonitrile in H_2O with 0.1% trifluoroacetic acid was used to elute the target peptide. The fraction containing the target peptide was lyophilized to provide a fluffy white powder (0.5 mg, 80% yield). The identity of the peptide was confirmed by MALDI-TOF mass spectrometry (*m/z* calculated for $C_{114}H_{162}N_{31}O_{31}$ 2462.7 $[M+H]^+$, found 2462.3 $[M+H]^+$). Peptide concentration (extinction coefficient of 22,300 $M^{-1}cm^{-1}$)was calculated from absorption at 280 nm using an extinction coefficient for Trp of 5575 $M^{-1}cm^{-1}$ [164].

2.2.5 Spectroscopy

Steady-state fluorescence spectra for all **tz** proteins (0.04 μM) were recorded on a QuantaMaster spectrofluorometer (Photon Technology International, London, Ontario) equipped with a Xenon arc lamp. Relative rates of trypsinolysis were determined with a Safire2 monochromator-based 96-well platereader (Tecan, Salzburg, Austria). To a solution of each purified FRET construct (0.08 μM) was added a buffered solution of trypsin to a final concentration of 0.5 μg/mL, and the fluorescence intensity at 530 and 480 nm (excitation at 430 nm) recorded every 5 s. All CD spectra were obtained with 8

μM protein in a 1 mm path length cuvette on an Olis DSM 17 CD spectrometer (Olis, Bogart, CA).

2.2.6 Description of the bacterial colony imaging system

The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dishes is a custom built device similar to one previously described [165]. The light from a 175 W xenon-arc lamp (Sutter Instrument Company, Novato, CA) is passed through a filter wheel (Sutter) holding either a 426 nm to 446 nm bandpass filter (Chroma Technology Corp., Rockingham, VT) or a 490 nm to 510 nm bandpass filter for excitation of CFP or YFP respectively. The filtered light then enters a bifurcated fiber optic bundle (Newport Corporation, Stratford, CT) that is positioned to illuminate a 10 cm dish placed in a recessed holder on the bench top. The fluorescence emission is filtered through a second filter wheel that is directly above the plate and that is holding either a 460 nm to 500 nm bandpass filter for CFP or a 520 nm to 550 nm bandpass filter for YFP. Images are acquired with a Retiga 1300i 12-bit CCD camera (QImaging, Burnaby, BC) positioned immediately behind the emission filter. With the exception of the lamp, the whole system is shrouded in black fabric to eliminate stray room light. Image Pro Plus (Media Cybernetics Inc., Silver Spring, MD) is used for automated camera control and digital image processing with custom macros. Typical exposure times are 1 s and all images are background corrected with identically acquired images of a Petri dish containing LB/agar but lacking bacterial colonies. Individual colonies are automatically identified within the digital images and the pixel intensities are summed in each of 3 images corresponding to the CFP channel, the FRET channel, and direct YFP channel. The measurement referred to as the YFP/CFP ratio in **Figure 20** and in **Figure 21** is the ratio of the average intensity in the FRET channel divided by the average intensity in the CFP channel for a single bacterial colony. The intensity of the YFP channel is directly proportional to the expression level of the fusion construct. For all imaging experiments, around 1000 bacterial colonies were grown on LB/agar supplemented with ampicillin (0.1 mg/mL) and arabinose (0.2%).

2.3 RESULTS AND DISCUSSION

2.3.1 Construction of CFP-peptide-YFP expression vectors

Table 1 Sequences of **tz** series peptides flanked by CFP and YFP

Protein	Peptide sequence[a]
tz1	KKWTWNPATGKWTWQE
tz2	KKWAWNPATGKWAWQE
tz3	KAWAWNPATGKWAWAE
tz4	KKWAWAPAAGAWAWQE
tz5	KAWAWAPATGAWAWAE
tz6	KKGTGNPATGKGTGQE

[a] Differences relative to **tz1** are underlined.

To validate our approach, we designed, expressed, and purified six proteins each consisting of an N-terminal CFP and a C-terminal YFP flanking an intervening peptide sequence provided in **Table 1**. The parent peptide in protein **tz1** is identical in sequence to the previously reported trpzip peptide **HP5W4**, which has been optimized for fold stability and reported to be >96 % folded at 298 K [155]. The peptide portions of **tz2** through **tz5** were designed to have varying degrees of destabilization due to the introduction of amino acids with low-sheet propensity [166] or low occurrence at particular hairpin positions [155]. The peptide portion of **tz6** has all four Trp replaced with Gly and was expected to be unstructured. The intrinsically ratiometeric design of the **tz** series CFP/YFP fusion proteins allowed us to obtain accurate and reproducible measurements of the FRET efficiencies by measuring the intensity of the donor CFP fluorescence before and after treatment with trypsin. Under the conditions of the proteolysis reaction, the peptide is cleaved and FRET is abolished, while the CFP remains intact. **Figure 16** shows the fluorescence spectra for proteins **tz1** and **tz6** before

and after proteolysis reaction. The FRET efficiency can be calculated from the CFP intensity before and after by equation $E = 1 - \frac{F_{DA}}{F_D}$ [162].

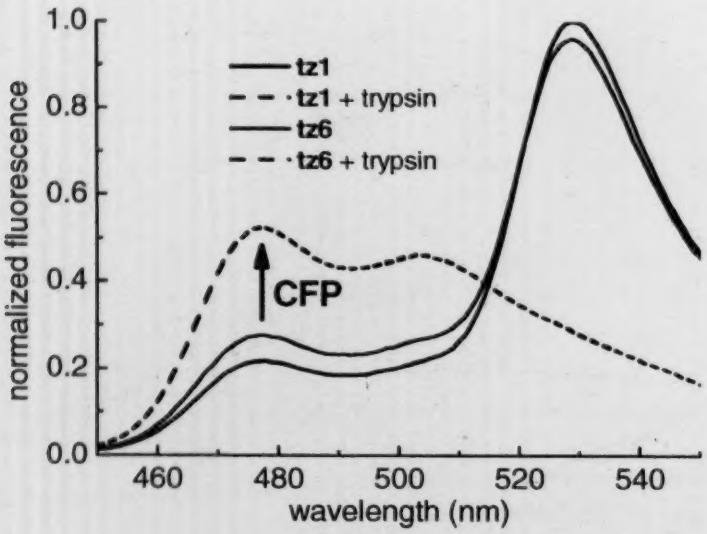


Figure 16 The corrected and normalized emission spectra of **tz1** and **tz6**. The spectra were corrected and normalized at 527 nm (excitation at 430 nm) from identical concentration of **tz1** and **tz6** before and after complete digestion of the fused peptide with trypsin.

2.3.2 *In vitro* characterization of tz series proteins

As shown in **Table 2**, the experimental FRET efficiencies determined for the **tz** series of proteins range from 47 % for **tz6** to 59 % for **tz1**. With the exception of **tz3** (discussed later), all proteins gave a distinct and characteristic FRET efficiency that correlated with the β-hairpin stability, as subsequently determined by orthogonal methods. The FRET efficiency of 59 % determined for **tz1** corresponds to a calculated interchromophore distance of 4.7 nm [55, 162]. This distance is consistent with molecular models of the full-length protein with a folded β-hairpin and the assumption of random chromophore

orientations. In support of this assumption, the **tz** series of proteins was designed with hydrophilic linkers of five residues in length connecting the inserted peptide to each of the flanking fluorescent proteins and with a CFP and YFP pair that does not dimerize [17]. The FRET efficiency of 47 % determined for **tz6** corresponds to an interchromophore distance of 5.1 nm. It is important to emphasize that the experimental FRET efficiencies are ensemble averages for populations of linked CFP and YFP pairs with a distribution of interchromophore distances [162]. CFP and YFP pairs with increased flexibility of their linkers probably have much broader distributions of interchromophore distances, yet the increase in the mean distance is apparently slight. Fortunately, in the **tz** series proteins, these small changes in the mean distance are at approximately the Förster distance (R_0 = 5 nm) [55], at which the corresponding change in FRET efficiency is maximal [162].

Table 2 FRET efficiencies, rates of proteolysis, and exciton strengths for proteins **tz1** through **tz6**

	FRET efficiency [%][a]	**Relative rate of trypsinolysis per lysine[b]**	**Relative exciton strength from CD[c]**
tz1	59.0 (±0.7)%[d]	1.00 (±0.08)[d]	1.0
tz2	55.5 (±1)%	2.39 (±0.2)	1.0
tz3	63.1 (±1)%	2.48 (±0.2)	0.80
tz4	51.4 (±2)%	18.5 (±2)	0.37
tz5	49.4 (±1)%	31.8 (±2)	0.04
tz6	47.2 (±1)%	14.6 (±0.4)	<0.01

[a] FRET efficiency (E) was determined from E=1-F_{DA}/F_D, where F_{DA} and F_D are the intensities of CFP before and after trypsinolysis, respectively. **[b]** Initial velocity of proteolysis divided by the number of lysine residues and normalized to **tz1**. **[c]** Baseline-corrected exciton strength in difference CD spectra relative to trypsinized **tz6**. **[d]** Standard deviation for ≥3 averaged measurements is reported as (±).

During the course of the FRET measurements, dramatic differences in the rates of trypsin digestion were noticed. A detailed investigation of the kinetics of proteolysis revealed that the rate of digestion is inversely correlated with the FRET efficiency (**Table 2** and **Figure 17**). The rates of proteolysis divided by the number of lysines in each peptide (1, 2, or 3) and normalized to **tz1** are shown in **Table 2**. Our interpretation of the observed trend is that highly folded peptides are poorer substrates for trypsin. This led us to conclude that both the FRET efficiency and rate of proteolysis are probing the structural stability of the β-hairpin peptide.

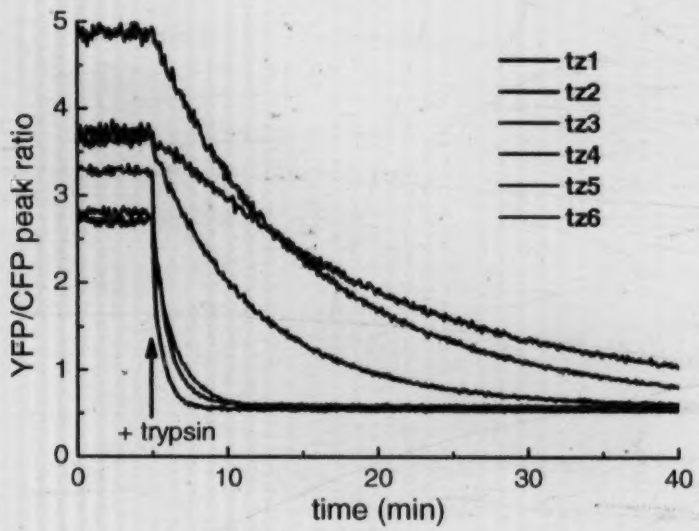


Figure 17 Ratio of YFP to CFP for purified proteins **tz1** through **tz6** upon treatment with trypsin. YFP fluorescence was recorded at 530 nm and CFP fluorescence at 480 nm (with excitation at 430 nm). The initial ratios are determined by the FRET efficiencies of the intact proteins. The final ratio (~0.5) corresponds to the 530/480 nm ratio of CFP itself and is stable over many hours demonstrating that the fluorescent protein is not digested under the conditions of this experiment.

To further characterize the β-hairpin structure in the context of the FPs, CD spectroscopy was employed. The far-UV CD spectrum of **tz1** is dominated by the primarily β-sheet

structures of CFP and YFP. However, trpzip peptides have a unique CD signature with strong bands at 215 and 229 nm due to exciton coupling of the indole side chains of Trp [146]. We have found that the exciton component is visible as a positive contribution at ~ 227 nm in the CD spectrum of **tz1** when compared to **tz6** (**Figure 18**).

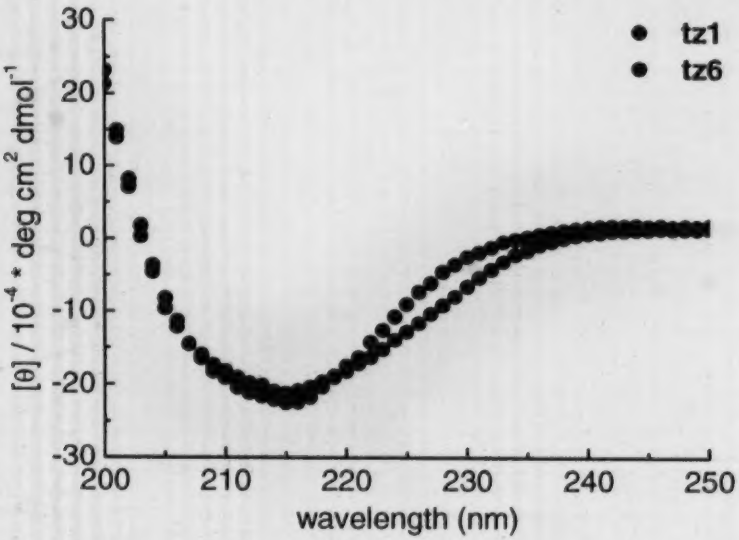


Figure 18 The far-UV CD spectra of identical concentrations of **tz1** and **tz6**

Subtraction of the CD spectrum of trypsinized **tz6** from the **tz1** spectrum clearly reveals the positive component of the Trp exciton (**Figure 19**). The negative component at 215 nm is poorly defined due to the strong contribution of the β-sheet structure at this wavelength. The exciton strengths (**Table 2**) provide a reliable measure of the relative amount of β-hairpin structure in the context of the full-length proteins. To assess the effect of the linked CFP and YFP proteins on the stability of the intervening β-hairpin structure, we measured the CD spectra for **tz1**-peptide, the peptide portion of **tz1** without linked CFP and YFP. As shown in **Figure 19**, the magnitude of the **tz1** exciton is 61 % of that measured for **tz1**-peptide; this demonstrates that the linked CFP and YFP cause only moderate destabilization of the intervening hairpin structure. This destabilizing effect

should be constant across the **tz** series of proteins and irrelevant to measurements of relative stability.

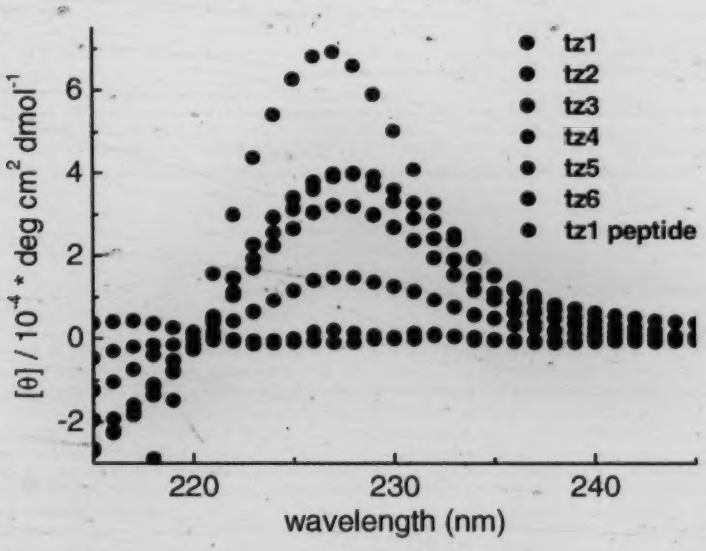


Figure 19 Difference CD spectra obtained by subtracting the CD spectrum of trypsinized **tz6** from each of **tz1** through **tz6**

2.3.3 *In vivo* characterization of tz series proteins

To assess whether FRET between CFP and YFP can reliably report on the structural stability of an intervening β-hairpin peptide *in vivo*, each of the **tz** proteins was expressed in *E. coli* and, by using a custom-built digital imaging system, the YFP/CFP ratio in individual colonies was determined. **Figure 20** shows a plot of the YFP/CFP emission ratio (excitation at 436 nm) *versus* YFP intensity (excitation at 500 nm) for six different populations of bacterial colonies, each of which is expressing a different member of the **tz** series of proteins. Each population exhibits a distinct and characteristic average ratio that is in qualitative agreement with the ranking of the *in vitro* FRET efficiencies provided in **Table 2**.

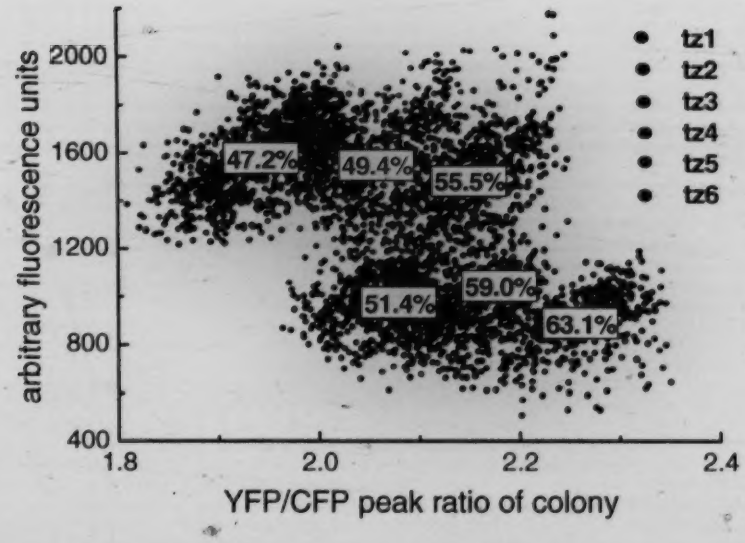


Figure 20 The YFP/CFP ratio of proteins **tz1** through **tz6** in live cells. Scatter plot of "in colony" YFP/CFP fluorescence ratio (excitation at 436 nm) against "in colony" YFP fluorescence intensity (excitation at 500 nm) extracted from digital images of hundreds of bacterial colonies (1 data point per colony) grown on LB/agar (1 Petri dish per **tz** protein). All plates were prepared and imaged under identical conditions. Superimposed on each set of data points is the *in vitro* FRET efficiency from **Table 2**. The fluorescence intensity of YFP (directly excited at 500 nm) is proportional to the amount of protein present in the colony and depends on colony age and camera settings.

It is interesting to note that the proteins with less structured peptides tended to accumulate at higher levels and that, within a given population, colonies with higher levels of protein tended to have higher ratios. To test whether protein aggregation was contributing to the observed differences in "in colony" ratio, we measured the ratio for proteins **tz1** and **tz6** in individual colonies that had accumulated intracellular protein concentrations of between 3 and 18 μM (**Figure 21**).

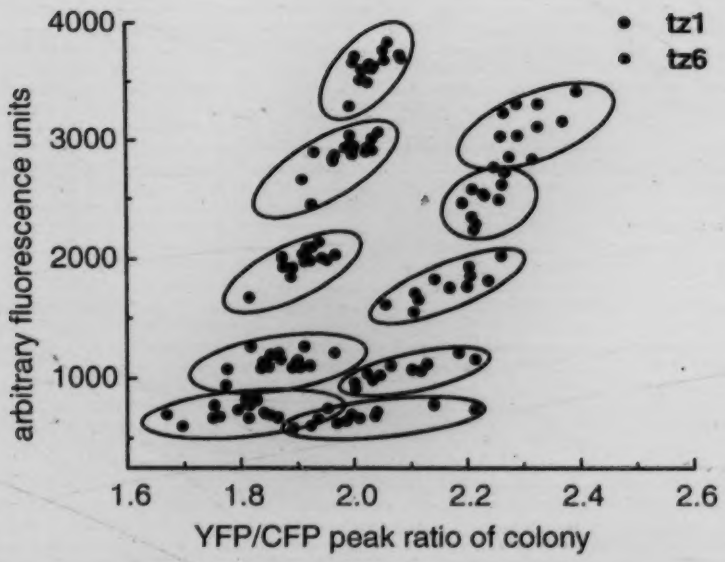


Figure 21 Relationship between the YFP/CFP peak ratio and the protein expression level. Shown in this chart are the YFP/CFP intensity ratios plotted against YFP fluorescence (direct excitation of YFP) for 16 individual bacterial colonies expressing **tz6** (red circles) and 11 individual colonies expressing **tz1** (black circles) imaged at 5 different time points at 20, 25, 35, 45, and 55 h after plating. Colonies in the same time point are grouped in cycles. By imaging similarly sized drops of purified protein under otherwise identical conditions we have arrived at an estimate of 1100 fluorescence units for 5 μM protein at 1 s exposure time. Correspondingly, the range of intracellular concentrations for the bacteria represented in this chart is approximately 3 to 18 μM.

Over this range of concentrations, colonies expressing either **tz1** or **tz6** exhibited a slight dependence of ratio on concentration; this suggested that intermolecular energy transfer was taking place. To investigate this possibility, we constructed and purified versions of the **tz1** protein in which either the YFP was replaced with CFP (CFP-**tz1**-CFP) or the CFP was replaced with YFP (YFP-**tz1**-YFP). We reasoned that if intermolecular associations were contributing to the observed YFP/CFP ratios, we should be able to observe FRET when CFP-**tz1**-CFP and YFP-**tz1**-YFP were mixed. Indeed, when we measured the emission spectra (excitation at 436 nm) for mixtures containing up to 15

μM of each protein, we did observe a concentration-dependent increase in the YFP emission peak (**Figure 22**).

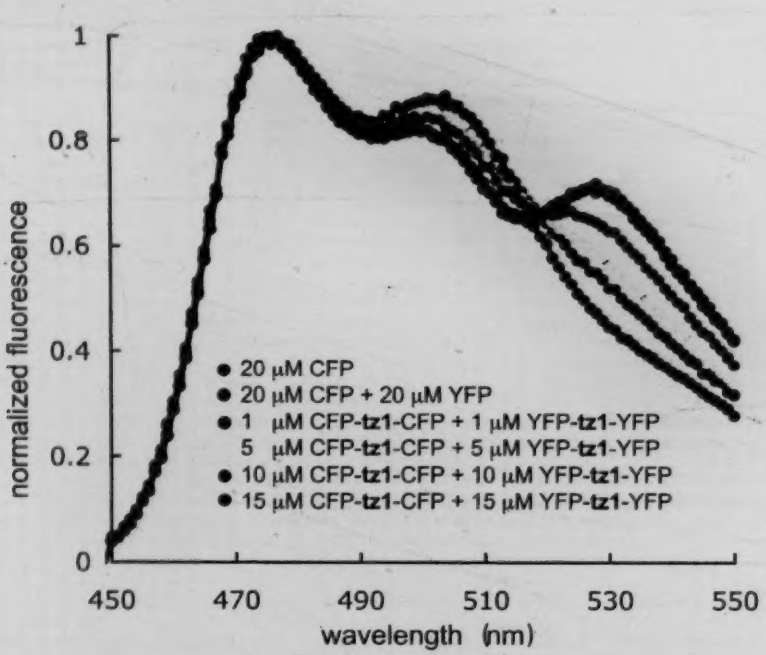


Figure 22 Concentration dependent fluorescent spectra. The **tz1** protein does not aggregate at typical intracellular concentrations. Shown in the chart are the *in vitro* fluorescence emission spectra of the indicated protein constructs. The apparent energy transfer peak at 530 nm is observed only at concentrations greater than 1 μM and is not diminished by treatment with trypsin. The magnitude of the effect and the preferential quenching of the longer-wavelength "hump" of the CFP emission is consistent with inner filtering of the CFP emission by the YFP chromophore.

However, this apparent energy transfer could not be abolished with trypsin and so could not be attributed to interactions between the peptide sequences. Rather, the concentration-dependent changes in YFP/CFP ratio observed both *in vitro* and *in vivo* are consistent with the inner filtering of the CFP emission by the YFP chromophore. Inner filtering was not observed at the concentrations (less than 100 nM) used for our FRET and trypsinolysis measurements. Our data strongly support the conclusion that the "in colony" ratios are determined by the amount of β-hairpin structure rather than potentially

confounding effects such as proteolysis or aggregation. A previous FRET-based approach for determining the stability of antibody V_L domains found that the dramatic differences between the range of FRET intensities observed *in vitro* (~10 %) and *in vivo* (~300 %) could only be explained by differences in proteolytic susceptibility [167].

2.4 CONCLUSION

The results described above have led us to conclude that the relative β-hairpin structural stabilities of the peptide portions of **tz1** through **tz6** are the same both *in vitro* and *in vivo* and decrease in the order **tz1>tz2>tz3>tz4>tz5>tz6**. This conclusion is strongly supported by the CD data, is reinforced by the proteolysis data, and is consistent with the fluorescence data. The FRET and proteolysis data sets are almost certainly biased by sequence-dependent factors, such as fluorescent protein orientation and trypsin substrate preference, which are irrelevant to fold stability yet difficult to account for. These factors could explain the anomalous ranking of **tz3** in the FRET data set and **tz6** in the proteolysis data set. While we must remain cognizant of these caveats, they do not undermine the fact that this approach is perhaps the most effective and versatile of the few reported methods for assessing the relative structural stability of a systematically modified series of peptides in live cells [168]. A corollary of these results is that genetic fusion of two relatively large proteins to the termini of a small peptide does not necessarily preclude proper folding of the peptide sequence.

Does the peptide portion of **tz1** qualify as a generic protein scaffold? Our CD results show that the presence of five destabilizing mutations, as in **tz4**, decreases the amount of folded structure by about half (**Table 2**). Since Ala has a less than average β-sheet propensity [166], we were probably overestimating the destabilizing effects of random substitutions. With this provision in mind, we proposed that the **tz1** peptide is an imperfect generic hairpin scaffold and that an average member of a library with random substitutions (except Pro) at positions 4, 6, 9, 11, and 13 (as in **tz4**) would be approximately 50 % folded. These results are extremely encouraging and have inspired

us to attempt development of a new β-hairpin variant with an improved tolerance for destabilizing mutations.

CHAPTER 3:

A RATIONAL APPROACH TO DEVELOP AN EXTENDED 20-MER β-HAIRPIN

The work of this chapter was published in as a research paper "Fluorescence-based characterization of genetically encoded peptides that fold in live cells: progress towards a generic hairpin scaffold", *Proceeding of SPIE*, 2007, **6649**: 66490S

3.1 INTROUDCTION

A desired generic protein scaffold for molecular recognition should be able to withstand multiple amino-acid substitutions/deletions/insertions without significant structural destabilization. We have proposed that trpzip might be a good candidate for a generic scaffold because of its remarkable thermal stability [1]. Unfortunately, when destabilizing amino acids presented, trpzip becomes significantly less structured. To circumvent this problem, an engineering strategy was undertaken to improve trpzip stability and will be discussed in this chapter. Generally, protein/peptide engineering falls into one of the following two approaches: combinatorial library screening and rational design. Both of which have their own merits and limitations.

The combinatorial approach consists of creating random mutants (libraries) of particular proteins and selecting variants with the desired properties. This strategy mimics the natural protein evolution, and numerous novel proteins have been engineered by this approach. Two of the major challenges in this method are creating "good" libraries and the need for efficient selection methods. With recombinant DNA techniques, trillions of variants could be obtained steadily. For example, 10 amino acids saturated mutagenesis in a protein using "NNK" degenerated codon will generate a size of $32^{10} = 1.1 \times 10^{15}$ library. But it is very difficult to produce that many variants or throughout screen each variant in such a big library for desired properties, even with today's sophisticated selection techniques. Therefore, a good library should have a balance between variability and its size. The methodologies to screen libraries efficiently are equally important. Although with the advent of different display techniques described in **Chapter 1**, the combinatorial approaches are still severely bottlenecked by the lack of high through-put selection methodologies.

Rational design, on the other hand, depends upon the detailed understanding of the targeted protein structure and/or other characteristic properties such as stability, function, and catalytic activities. Rational design is a precise protein mutagenesis technique rather than requiring complicate processes to extensively screen individual variant in large libraries. It usually provides a simple approach to alternate protein properties, though it does not always generate positive results. Therefore, advanced computer programs are under development to facilitate rational protein design [169].

Nevertheless, considering the simplicity of the trpzip peptides structure and its relatively small size (16 amino acids); on top of which detailed structures are available, rational design might be an efficient approach in engineering a more stable trpzip peptide, which perhaps would be a good candidate to be a generic β-hairpin scaffold. As described in **Chapter 2**, by comparing CD, trypsinolysis, *in vitro* and *in vivo* FRET data, we concluded that trpzip (**tz1**) is an imperfect candidate to be a generic protein scaffold, because it only remains 50% folded when destabilizing substitutions are present at positions 4, 6, 9, 11, and 13 [1]. Trpzip stability arises from the $\pi - \pi$ interaction in the two interdigitating cross-strand tryptophan-tryptophan pairs [146]. Given the generally favourable properties of the trpzip-type peptides, we hypothesized that an elongated β-hairpin with a 3rd cross-strand Trp/Trp pair could potentially impose more structural constraint to the overall structure, presumably it will form a more stable β-hairpin peptide. We therefore designed a new peptide, designated extended trpzip2 (**xtz2**) (**Table 3**). **xtz2** is a 20mer β-hairpin that is "extended" by four residues relative to its **tz1** counterpart. The four additional residues in **xtz2** are a Trp/Thr pair inserted between the 2nd and 3rd residues and a Thr/Trp pair inserted between the 14th and 15th residues of the **tz1** β-hairpin (**Figure 23**).

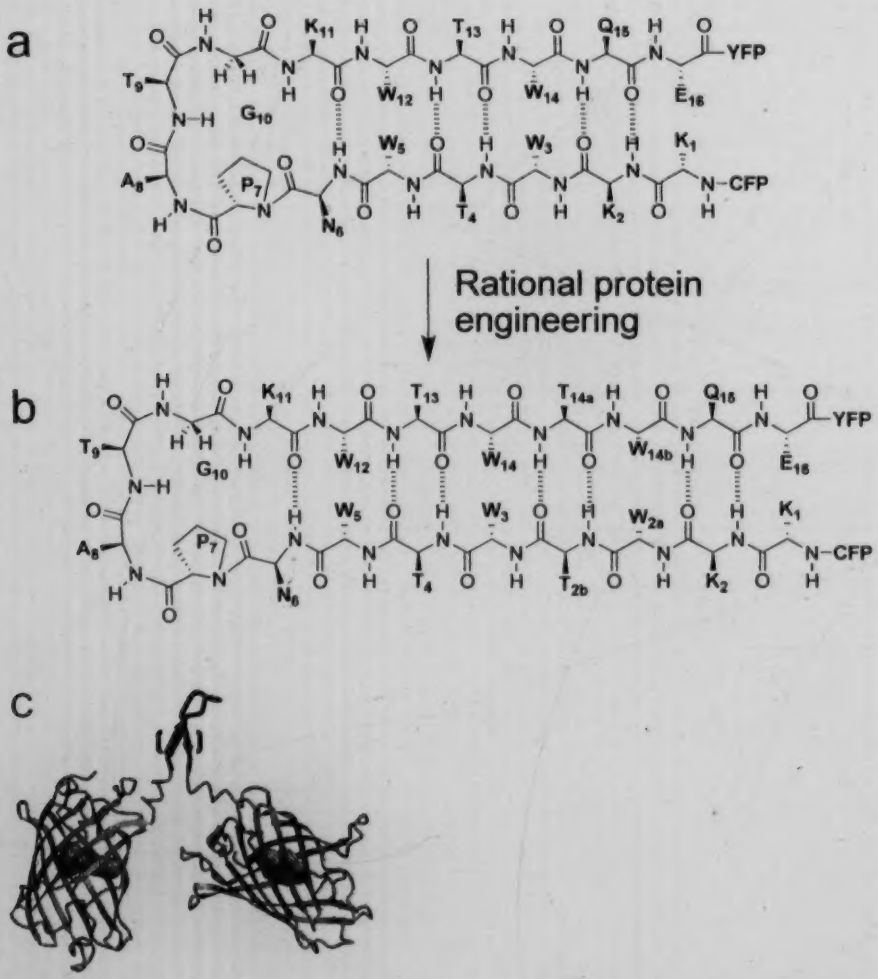


Figure 23 Representations of the β-hairpin peptide in the context of the CFP/YFP fusion protein: **(A) tz1** and **(B) xtz2**. **(C)** "To-scale" cartoon of a complete FRET construct with the location of inserted residue pairs indicated.

In analogy to the **tz** series proteins [1], we constructed three additional variants of **xtz2**: one unstructured control peptide lacking tryptophan (**xtz3**); one with 5 destabilizing substitutions (**xtz4**); and one with 6 destabilizing substitutions (**xtz5**). The sequence of the peptide portions of the **xtz** series proteins is provided in **Table 3**.

Table 3 Sequences of **xtz** series peptides flanked by CFP and YFP

Protein	Peptide sequence[a]
tz1	KK--WTWNPATGKWTW--QE
xtz2	KKWTWTWNPATGKWTWTWQE
xtz3	KKGTGTGNPATGKGTGTGQE
xtz4	KKWTWAWAPAAGAWAWTWQE
xtz5	KKWAWAWAPATGAWAWAWQE

[a] Differences relative to **tz1** are underlined.

3.2 MATERIAL AND METHODS

3.2.1 General procedures and materials

Please refer to **Section 2.2.1**.

3.2.2 Constructions of CFP-xtz2-YFP and CFP-xtz2-peptide expression vectors

The CFP-**xtz**-YFP vector was constructed based on the CFP-peptide-YFP vector described in **Section 2.2.2**. Briefly, dsDNA encoding individual **xtz2** series sequence with appropriate sticky ends were created by slow cooling pairs of complementary single stranded oligonucleotides from 95 °C to room temperature. This synthetic annealed dsDNA was ligated into the CFP-peptide-YFP expression vectors that had been digested with *Sac*I and *Kpn*I. Ligated product was used to transform *E. coli* DH10B (Invitrogen) by electroporation. The **xtz2** series proteins were expressed under the same conditions as described in **Section 2.2.4**. The amount of soluble **xtz2** series proteins in bacteria

decreases noticeably while comparing with **tz1** series. It may be attributed to the extra Trp-Trp pair in the **xtz2** series, which is not favorable for protein folding in bacteria.

The CFP-**xtz2**-peptide expression vector was constructed from the CFP-**tz1**-peptide expression vector described in **Section 2.2.3**.

3.2.3 Protein and peptide purification

FRET constructs were purified as described in **Section 2.2.4**. The **HP5W4** peptide was purchased from Sigma-Genosys Canada. The **xtz2**-peptide was expressed and isolated using a previously described bacterial expression system [1]. All peptides were further purified by reversed-phase chromatography on a Prosphere HP C18 300A column (Alltech Associate, Inc.). A linear gradient of increasing acetonitrile in H_2O with 0.1% trifluoroacetic acid was used to elute the target peptides. The fraction containing the target peptide was lyophilized to provide a fluffy white powder. Calculated extinction coefficients of 22,300 $M^{-1}cm^{-1}$ and 33,450 $M^{-1}cm^{-1}$ at 280 nm were used to determine the concentration of **HP5W4** and **xtz2**-peptide respectively.

3.2.4 Spectroscopy

Please refer to **Section 2.2.5**.

3.3 RESULTS AND DISCUSSION

3.3.1 *In vitro* characterization of xtz2 series proteins

Proteins **xtz2**, **xtz3**, **xtz4**, and **xtz5** were expressed in *E. coli* and purified by a procedure identical to that previously reported for the **tz** series proteins (**Chapter 2**) [1]. After lysing the bacteria, except **xtz3**, a high percentage of the recombinant proteins aggregated

in the bacterial inclusion body was found. This prevented the proteins isolation by Ni-NTA affinity purification. The yields of purified **xtz** series proteins were typically ~30% of their **tz**-series counterparts.

Table 4 FRET efficiencies and rates of proteolysis for proteins **tz1** and **xtz2** through **xtz5**

Protein	FRET efficiency (%)	Relative rate of trypsinolysis per K/R[a]
tz1	59.0 (±0.7)%	1.00 (±0.08)[b]
xtz2	61.7 (±1.2)%	0.14 (±0.01)
xtz3	44.0 (±0.8)%	12.8 (±0.6)
xtz4	58.3 (±0.7)%	3.26 (±0.9)
xtz5	53.1 (±0.4)%	8.76 (±0.2)

[a] Rates are the initial velocities of trypsinolysis divided by the total number of lysine residues in the peptide sequence, relative to tz1 which was assigned a value of 1.00. [**b**] Errors for three or more independent measurements are reported in parenthesis as ± standard deviation.

As shown in **Table 4**, the experimentally determined FRET efficiency was highest for the **xtz2** protein containing the non-destabilized parent peptide and decreased in the order (highest FRET) **xtz2** > **xtz4** > **xtz5** > **xtz3** (lowest FRET). The rates of proteolysis relative to **tz1** were inversely correlated with FRET efficiency (**Table 4** and **Figure 24**), increasing in the order (slowest proteolysis) **xtz2** < **xtz4** < **xtz5** < **xtz3** (fastest proteolysis). By analogy with the **tz** series, which contained variants with similar destabilizing mutations [1], we had expected a stability ranking of (most stable) **xtz2** > **xtz4** > **xtz5** > **xtz3** (unstructured). Indeed, this ordering is consistent with both the FRET and resistance to proteolysis data.

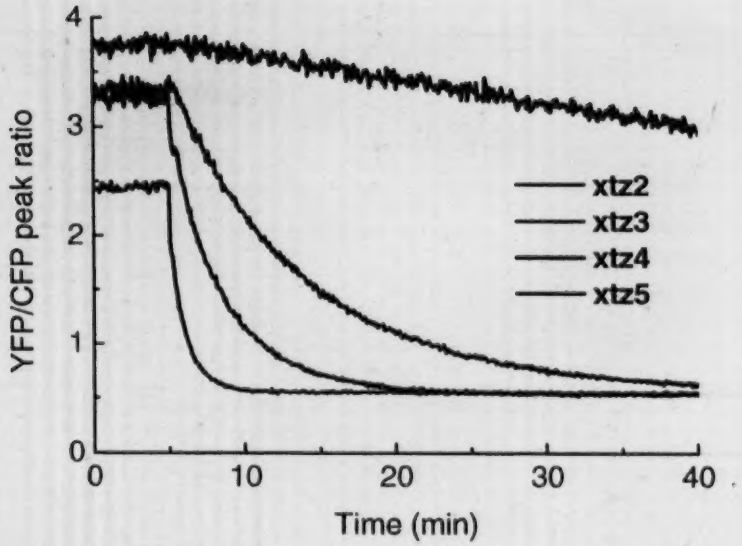


Figure 24 Ratio of YFP to CFP for purified proteins **xtz2** through **xtz5** upon treatment with trypsin. Trypsinolysis cleaves the acceptor YFP from the donor CFP and results in a loss of FRET. The final ratio (~0.5 in all cases) corresponds to the donor CFP only and is stable over many hours indicating that the CFP portion of the protein is not susceptible to trypsinolysis under the conditions of the experiment. The relative initial rates of proteolysis are provided in **Table 4**.

Trpzip-type peptides have strong bands in their CD spectra at 215 and 229 nm due to exciton coupling of the two pairs of edge-to-face packed indole side chains of tryptophan [146, 170]. In **Chapter 2**, a method was described to measure the tryptophan exciton by subtracting the **tz2** CD spectrum from different **tz** series protein CD spectra. If the **xtz2** series peptides have similar β-hairpin structures as **tz1** series when fused between two FPs, they are expected to produce similar CD spectra after analogous spectrum subtraction. Subtraction of the spectrum of trypsinized **xtz3**, which lacks the four tryptophan residues and is expected to be completely unstructured, from the spectrum of intact **xtz2** unmasks the expected exciton couplet in the difference CD spectrum (**Figure 25**). Likewise, the CD spectrum of trypsinized **xtz3** was subtracted from those of proteins **xtz4** and **xtz5**. The strength of the exciton couplet decreased in the order (strongest

couplet) **xtz2** > **xtz4** ~ **xtz5** (weakest couplet). The relative strengths of the exciton couplets reflects the relatively fold stability of the β-hairpin peptides as described in Chapter 2. By comparing the change in FRET efficiency, the rate of proteolysis and the tryptophan CD exciton of both **tz1** and **xtz2** series, we can conclude that the **xtz2** peptide retains a relatively stable structure, even when five or six destabilizing substitutions are present.

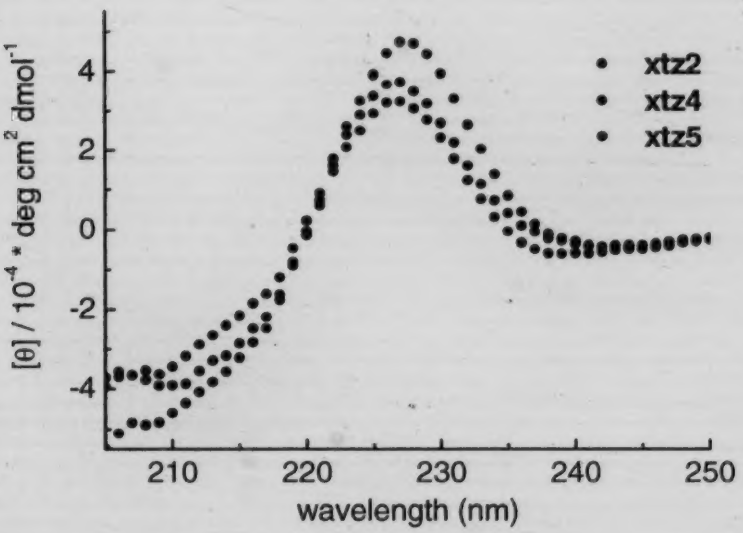


Figure 25 Difference CD spectra obtained by subtracting the CD spectrum of trypsinized **xtz3** from each of intact **xtz2**, **xtz4**, and **xtz5**.

3.3.2 *In vivo* characterization of xtz2 series proteins

The *in vitro* characterization of the **xtz** series of proteins provided strong support for the β-hairpin portion of **xtz2** being highly structured, and the β-hairpin portions of **xtz4** and **xtz5** somewhat being less so. We next asked whether this trend would also be true of these peptides when presented in the intracellular milieu. To answer this question, we expressed each **xtz** series protein in *E. coli* and, using a custom-built digital imaging

system, we determined the ratio of YFP to CFP fluorescence (excitation of CFP at 430 nm) in individual colonies. This approach does not provide the actual FRET efficiency but it does provide a YFP/CFP intensity ratio that is proportional to the FRET efficiency. In practice, we found that it is difficult to obtain a perfect correlation between the "in colony" ratios and the *in vitro* FRET efficiencies, apparently because of an age-dependent colony autofluorescence and inner filter effects that we have been unable to properly correct for (**Chapter 2**) [1]. However, if the bacteria are plated at a consistent density (100-250 per plate), and all plates are treated in an identical fashion, ratios measured on different plates display excellent qualitative agreement with the *in vitro* FRET efficiencies.

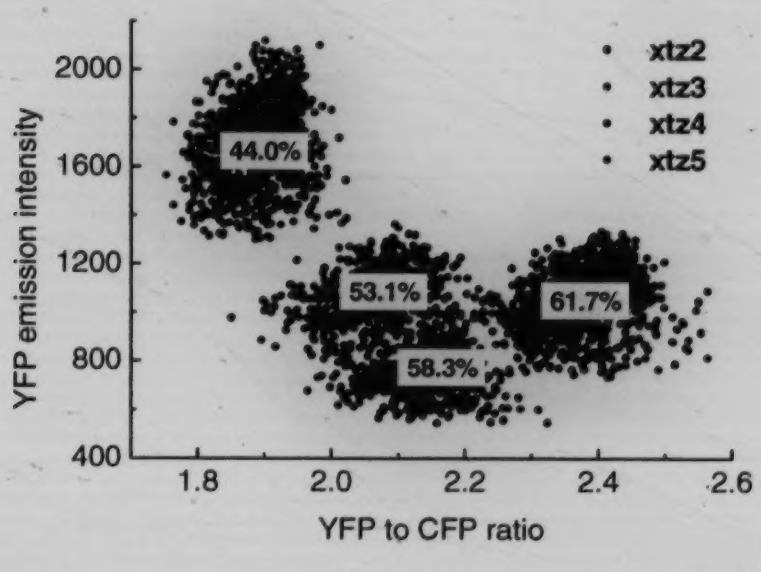


Figure 26 The YFP/CFP ratio of proteins **xtz2** through **xtz5** in live cells. The YFP/CFP peak ratio for each individual colony was determined by ratioing the average intensity in the YFP channel by the average intensity in the CFP channel. The YFP intensity for each individual colony is the average intensity for YFP emission when YFP is directly excited at 500 nm. Superimposed on each set of data points is the *in vitro* determined FRET efficiency for the purified protein.

Shown in **Figure 26** is a representative ratio of data obtained for **xtz2** through **xtz5** proteins on separate plates and imaged "in colony". We have plotted these data as colony ratio *vs.* colony intensity to better represent the intrinsic variability observed for these bacterial homogeneous populations. We attribute the pronounced and reproducible differences in colony intensity to protein-specific differences in expression levels. From these results we conclude that the relative stability of the **xtz** series β-hairpins in the intracellular environment is similar to their *in vitro* stability.

3.3.3 Characterization of the isolated peptide portion of xtz2

Encouraged by the apparent success at designing a highly stable extended β-hairpin structure, we recombinantly expressed and purified the peptide portion of **xtz2** (designated **xtz2**-peptide) with no fluorescent proteins attached to either end. The **tz1**-peptide (**HP5W4**) was purchased from Sigma-Genosys Canada. Trpzip type β-hairpin peptides are typically soluble up to at least millimolar concentrations in buffered aqueous solution and show the characteristic exciton couplet at 215 nm and 227 nm in their CD spectrum [146]. Since the **xtz2**-peptide has significantly more exposed hydrophobic surface than other previously reported trpzip-type peptides, the fact that it was not soluble at concentrations greater than ~70 μM did not come as a great surprise. However, we were quite surprised to find that the CD spectrum of this relatively dilute peptide showed no peak attributable to the exciton couplet (**Figure 27**). Addition of 50% trifluoroethanol (TFE) slightly improved the solubility and induced the formation of the trpzip-type β-hairpin structure as revealed by the appearance of the characteristic exciton couplet in the CD spectrum. It is apparent that the **xtz2**-peptide in aqueous solution only folds into a stable β-hairpin structure when in the context of the full-length **xtz2** fusion protein with CFP fused on one end and YFP on the other. So, despite its excellent tolerance of multiple destabilizing mutations, the poor solubility and context-dependent folding of the **xtz2**-peptide eliminates it as a candidate generic protein scaffold.

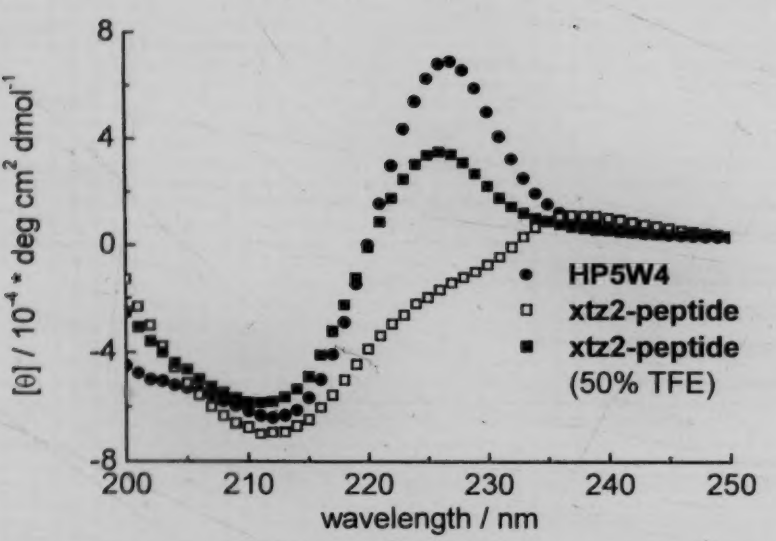


Figure 27 CD spectra of identical concentration of **HP5W4**, **xtz2**-peptide in buffer and **xtz2**-peptide in 50% TFE. The strong peak at 227 nm is a characteristic feature of trpzip-type peptides and is attribute to exciton coupling of the closely packed tryptophan residues [146].

3.4 CONCLUSION

The β-hairpin peptide portion of **xtz2** exhibits some, but not all, of the properties desirable in a generic protein scaffold. This 20-mer β-hairpin could tolerate as many as six destabilizing mutations without significant loss of secondary structure, but once removed from the context of the CFP and YFP fusion protein, it exhibited very poor solubility and did not form the expected β-hairpin structure in the absence of TFE. The ideal β-hairpin scaffolds should be highly soluble, exhibit efficient and context-independent folding, and have a high tolerance for destabilizing mutations. It is clear that our rational protein design strategy did not provide us a satisfactory candidate. To reach our goal, a combination of the *in vivo* FRET-based screening and the *in vitro* proteolysis

assay was be employed to select highly stable β-hairpins from large libraries expressed in live cells (as discussed in **Chapter 4**).

CHAPTER 4:

IN VIVO SCREENING IDENTIFIES A HIGHLY FOLDED β-HAIRPIN PEPITDE WITH A STRUCTURED EXTENSION

The work of this chapter was published as a research paper "In vivo screening identifies a highly fold beta-hairpin peptide with a structured extension", *Chembiochem*, 2007, **8**: 880-883.

4.1 INTRODUCTION

As described in the **Chapter 3**, we had attempted to engineer a more stable "extended" trpzip variant **xtz2** by rational design. This peptide showed some, but not all the desirable properties to be a generic scaffold: **xtz2** peptide has a greater stability in tolerating destabilizing substitutions as compared to its precursor **tz1**. However, its poor solubility is problematic for intracellular applications [2, 3]. It is clear that rational protein design shows less competence in trpzip stability engineering. Extensive combinatorial library design coupled with through-put screening provides an alternative approach in protein engineering [171].

The conventional approach in developing folded peptides involves chemical synthesis of systematically-modified peptide variants and individual characterization by CD and/or NMR spectroscopy [146, 152-155]. The use of synthetic peptides and reliance on low-throughput characterization techniques necessarily restricts the sequence diversity that can be explored, though efforts have been made to overcome this limitation [156, 172, 173]. In **Chapter 2**, an approach was described that enabled us to assess the ability of peptides to fold into β-hairpins in the cytoplasm of live cells [1]. Our strategy entails the recombinant expression of a peptide gene fused in frame between flanking genes encoding a CFP and a YFP. If a particular peptide sequence adopts a folded structure, CFP and YFP are brought into closer proximity and exhibit a higher efficiency of FRET. Higher FRET efficiency enhances the YFP (acceptor) fluorescence at the expense of the CFP (donor) fluorescence. Imaging of plates harbouring colonies of transformed bacteria provides the YFP and CFP fluorescence emission intensities for each colony, and thus peptides that are highly folded *in vivo* can be distinguished from those that are not.

As will be described later in this chapter, this FRET-based approach also provides a versatile method for screening large libraries of peptide sequences for highly structured variants. We have applied this strategy to the development of a version of trpzip-type β-hairpin [146] with a structured extension. The minimal and highly stable trpzip structure has emerged as a preferred model system for computational and experimental studies of protein folding [174-176]. We have proposed that trpzip-type peptides (or tandem fusions of such peptides) could serve as a minimal protein scaffold for molecular recognition in the cytoplasm of live cells [1, 3]. With this goal in mind, we sought to employ a library-based screening strategy to identify a candidate trpzip-type peptide with high fold stability *in vivo*. A similar approach has previously been used to increase the thermal stability of engineered immunoglobulin V_L domains [167].

4.2 MATERIAL AND METHODS

4.2.1 General procedures and materials

Please refer to **Section 2.2.1**.

4.2.2 Library construction and screening

The CFP-peptide-YFP libraries were expressed using the previously described vector pZC1 in **Section 2.2.2**. This vector can be used to encode proteins of the general structure N_{term}-His6-EK-CFP(1-230)-TSGAQ-peptide-GTSAEYFP-YFP(5-238)-C_{term}. In **Chapter 2**, "peptide" represented a 16-mer from the **tz** series. In this chapter, "peptide" represents either a peptide library or an individual member of the **xtz** series. To create the 20-mer peptide libraries, dsDNA encoding the peptide libraries with appropriate sticky ends were created by slow cooling pairs of complementary single stranded oligonucleotides from 95 °C to room temperature. For the 24-mer and 28-mer peptides libraries, the dsDNA encoding the desired libraries were created by PCR amplification followed by *Sac*I and

*Kpn*1 restriction enzymes digestion. The resulted dsDNA, either synthetic annealed DNA or PCR product, was ligated into the CFP-peptide-YFP expression vectors that had been digested with *Sac*I and *Kpn*I. Ligated product was used to transform *E. coli* DH10B (Invitrogen) by electroporation. The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dish was described previously in detail in **Section 2.2.6**. Individual colonies were automatically identified within the digital images and the pixel intensities are summed in each of 3 images corresponding to the CFP channel, the FRET channel, and direct YFP channel. For each colony the ratio of YFP to CFP fluorescence emission intensity is calculated by dividing the average intensity in the FRET channel by the average intensity in the CFP channel. Colonies that had the highest ratio of YFP to CFP emission and the highest brightness were picked from the plate and cultured as described in **Section 2.2.4**. The sequence of the peptide portion of all selected proteins was determined by DNA sequencing with a forward primer (5'-CCCTCGTGACCACCCTGACCTGG-3') that anneals to the chromophore region of the gene for CFP.

4.2.3 Protein and peptide purification

FRET constructs were purified as described in **Section 2.2.4**. The isolated peptide portions of **tz1** and **xtz1**, are named **HP5W4** and **xtz1**-peptide respectively, were purchased from Sigma-Genosys Canada. The **xtz2**-peptide was isolated using a previously described bacterial expression system [1]. All peptides were further purified by reversed-phase chromatography on a Prosphere HP C18 300A column (Alltech Associate, Inc.). A linear gradient of increasing acetonitrile in H_2O with 0.1% TFA was used to elute the target peptides. The fraction containing the target peptide was lyophilized to provide a fluffy white powder. Calculated extinction coefficients of 22,300 $M^{-1}cm^{-1}$, 27,875 $M^{-1}cm^{-1}$ and 33,450 $M^{-1}cm^{-1}$ at 280 nm were used to determine the concentration of **HP5W4**, **xtz1**-peptide and **xtz2**-peptide, respectively.

4.2.4 Spectroscopy

Please refer to **Section 2.2.5**.

4.2.5 Protein expression and imaging in HeLa cells

To create the CFP-peptide-YFP mammalian expression plasmids, restriction sites encoding an *Xho*I and an *EcoR*I restriction site were introduced into pcDNA3 (Invitrogen) by a PCR-based method. The pBAD/His B plasmid containing the gene encoding CFP-peptide-YFP was digested with *Xho*I and *EcoR*I and ligated into the modified pcDNA3 plasmid digested with the same two enzymes. All DNA for mammalian cell transfection was purified by Plasmid Midi kit (Qiagen). HeLa cells were cultured at 37 °C in DMEM (Dulbecco's modified Eagle's medium; Invitrogen) supplemented with 10% (v/v) FBS (Sigma). Cells in 35 mm imaging dishes were transfected with 4 μg plasmid DNA mixed with 10 μg poly(ethylenimine) (linear, molecular weight ~25,000, Polysciences Inc.) in 0.5 ml OptiMEM (Invitrogen) and FBS was added after 3 hours. Approximately 14 to 24 h later, the medium was exchanged for Hanks' Balanced Salt solution (HBSS) containing no calcium chloride, magnesium chloride, magnesium sulfate, or phenol red (Invitrogen). Cells were imaged with a Zeiss Axiovert 200M epi-fluorescence inverted microscope (Zeiss) equipped with a monochrome Retiga 2000R 12-bit cooled CCD camera (QImaging). The light from a xenon arc lamp is passed through an external filter wheel (Sutter) holding either a 426 nm to 446 nm bandpass filter (Chroma Technology Corp.) or a 490 nm to 510 nm bandpass filter for excitation of CFP or YFP respectively. The fluorescence emission was passed through a second filter wheel holding either a 460 nm to 500 nm bandpass filter or a 515 nm to 555 nm bandpass filter for CFP or YFP respectively.

FRET efficiencies were obtained by measuring the intensity of the CFP signal (direct excitation) before and after bleaching of the YFP acceptor. FRET efficiencies reported in **Figure 35** represent the average of at least 15 independent single-cell measurements.

4.2.6 NMR spectroscopy and structure calculation

Synthetic **xtz1**-peptide (Sigma-Genosys) was dissolved in 90% H_2O / 10% D_2O to a final concentration of 2 mM. The pH was determined to be between 5 and 6. To this solution was added 0.010 ml of 0.1% the chemical shift reference 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). All chemical shifts are quoted relative to the DSS reference set at 0.00 ppm.

Spectra were recorded on a Varian Inova 600MHz spectrometer. Data acquisition and initial Fourier transformations were performed using VNMR 6.1C. Two-dimensional (2D) total correlation spectroscopy (TOCSY) [177, 178], gradient enhanced double quantum filter correlation spectroscopy (gDQF-COSY) [179], and nuclear Overhauser enhancement spectroscopy (NOESY) [180] were carried out. Solvent suppression was achieved by presaturation of the water peak. Spectral widths were set to be equivalent in both dimensions for all 2D spectra. All spectra were acquired in the phase sensitive mode using the States-Haberkorn-Ruben method [181], with the exception of the gDQF-COSY spectrum which was acquired in the absolute value mode selecting N-type signals. The gDQF-COSY and NOESY spectra employed 8192 data points in the directly detected dimension, whereas the TOCSY used 4096 data points in the directly detected dimension. The NOESY spectrum was recorded with 512 experiments in the indirectly detected dimension, and 32 scans were acquired per experiment, whereas the TOCSY and gDQF-COSY experiments were acquired with 256 experiments in the indirectly detected dimension with 16 scans per experiment. All spectra employed squared sine bell and shifted squared sine bell functions as weighting functions in both dimensions. The TOCSY experiments employed a spin lock mixing time of 100 ms and the mixing time for the NOESY experiment was 150 ms.

The choice of temperature for all experiments was based on a series of one-dimensional spectra acquired at 5 °C, 15 °C, and 27 °C. The amide region of these spectra was examined to determine at which temperature contained the least spectral overlap for all

amide resonances. 15 °C was found to be the most suitable temperature. $J_{HN-H\alpha}$ coupling constants were obtained from the 1D spectrum at 15 °C where possible.

Spectra were viewed in NMRview v.6.4 [182] and chemical shift assignments (**Appendix Table A1**) were manually determined. Automated NOESY assignment and structure ensemble calculation was performed with CYANA v.2.1 [183, 184]. A total of 431 unassigned NOESY peaks with intensity greater than ~1% of the most intense NOESY peak were input to CYANA. At less than 1% of the most intense NOESY peak, spectral noise was significant and peak assignments were unreliable. CYANA assigned 403 of 431 (93.5%) peaks, of which 309 were used as distance constraints in the structure calculation (**Appendix Table A2**).

4.3 RESULTS AND DISCUSSION

4.3.1 An "irrational" approach to develop an extended and highly folded 20-mer β-Hairpin

The template for our initial "extended trpzip" library was a 20-mer version of the highly folded 16-mer trpzip **HP5W4** [147]. The 20-mer contained two additional pairs of residues, one random and one threonine, genetically inserted after what would otherwise have been the second and 14th residues of **HP5W4**. The sequence of the resulting "peptide" portion of the 400-member protein library was $K^1\ K^2\underline{X}^3\underline{T}^4W^5T^6W^7N^8P^9\ A^{10}T^{11}G^{12}\ K^{13}W^{14}T^{15}W^{16}\underline{T}^{17}\underline{X}^{18}Q^{19}E^{20}$; here X represents all 20 amino acids (using degenerate codon "NNK"), and the residues inserted relative to **HP5W4** are underlined. Assuming β-strand conformation, the randomized positions would be directed towards the face of **HP5W4** that harbors the interdigitated Trp side chains.

Table 5 Peptide sequences identified by library screening in **Chapter 4**

	Library[a] Sequence	Sequences identified[b]			
		X^3	X^{18}		#
1	$K^1K^2X^3T^4...T^{17}X^{18}Q^{19}E^{20}$	W	R		3
		W	K		2
		X^2	$(K/R)^{18}$	X^{19}	#
2	$K^1X^2W^3T^4...T^{17}(K/R)^{18}X^{19}E^{20}$	A	K	N	2
		S	K	N	2
		P	K	N	1
		Q	R	N	1
		$(A/S)^2$	X^4	X^{17}	#
3	$K^1(A/S)^2W^3X^4...X^{17}K^{18}N^{19}E^{20}$	A	T	R	3[c]
		A	S	K	2
		A	S	R	1
		A	S	V	1

[a] "..." denotes residues 5 to 16 of the peptide. Positions marked "X" were subject to saturation mutagenesis by using the degenerate codon "NNK" where N = adenine (A), guanine (G), cytosine (C), or thymine (T), and K = G or T. In round 2, position 18 was mutated to either Lys or Arg by using the codon "ARG", where R = A and G. In round 3, position 2 was mutated to either Ala or Ser by using the codon "KCC". [b] Peptide sequence determined by DNA sequencing of plasmid DNA. [c] This sequence was designated as **xtz1**.

E. coli was transformed with the gene library, and ~6×10^3 colonies on 10 Petri dishes were subjected to fluorescence imaging in order to identify those exhibiting the highest ratio of YFP-to-CFP fluorescence emission. In **Chapter 2**, the ability of this imaging system to reliably distinguish colonies that express FRET constructs of various FRET efficiencies was demonstrated. The fluorescent brightness of each colony, considered commensurate with peptide solubility [185], was determined by direct excitation and

imaging of YFP. Five individual colonies that exhibited both a high ratio of YFP-to-CFP fluorescence emission and high brightness were cultured overnight, and the plasmid DNA was purified. DNA sequencing revealed a striking consensus at the randomized positions: in all five sequences, position 3 was Trp and position 18 was either Lys or Arg (**Table 5**). This consensus sequence was used as the new template for a second library in which positions 2 and 19 were similarly randomized. Library screening and DNA sequencing revealed a strong preference for Asn at position 19 and a weaker preference for small side chains, such as Ala and Ser, at position 2. An iterative third library with randomization of positions 4 and 17 was constructed and screened as described above. The CFP-YFP fusion protein containing the sequence identified most often in this third library was designated extended trpzip1 (**xtz1**) and subjected to further investigation.

4.3.2 *In vitro* characterization of xtz1 series proteins

Table 6 Summary of proteins and peptides compared in **Chapter 4**

Protein name	Peptide name	Peptide sequence (flanked by CFP and YFP in protein)
tz1	**HP5W4**	KK--WTWNPATGKWTW--QE
xtz1	**xtz1**-peptide	KAWTWTWNPATGKWTWRKNE
xtz2	**xtz2**-peptide	KKWTWTWNPATGKWTWTWQE
xtz3	n/a	KKGTGTGNPATGKGTGTGQE
xtz6	n/a	KAWTWAWAPAAGAWAWRKNE
xtz7	n/a	KAWAWAWAPATGAWAWAKNE

In analogy to the **tz** and **xtz2** series proteins, two additional variants of **xtz1** were constructed: one with 5 destabilizing substitutions (**xtz6**); and one with 6 destabilizing substitutions (**xtz7**). The sequence of the peptide portions of the **xtz1** series proteins are

provided in **Table 6**. Three additional CFP-YFP fusion proteins were investigated as comparisons and controls (**Table 6**). The first of these has been previously described as **tz1** [1] which contains a 16mer peptide identical to **HP5W4** [155]. The second contains a 20mer peptide with a potential third Trp/Trp cross-strand pair (**xtz2**, equivalent to **xtz1** with $K^2W^3T^4$ and $T^{17}W^{18}Q^{19}$). The third is a control protein that contains an unstructured 20mer peptide (**xtz3**, equivalent to **xtz2** with all six Trp replaced with Gly).

Table 7 FRET efficiencies and rates of proteolysis for proteins **tz1**, **xtz1** through **xtz7**.

Protein	FRET efficiency (%)	Relative rate of trypsinolysis per K/R[a]
tz1	59.0 (±0.7)%	1.00 (±0.08)[b]
xtz1	67.9 (±1)%	0.29 (±0.02)
xtz2	61.7 (±1)%	0.14 (±0.01)
xtz3	44.0 (±0.8)%	12.8 (± 0.6)
xtz6	63.2 (±0.2)%	6.19 (±0.1)
xtz7	60.3 (±1)%	7.85 (±0.1)

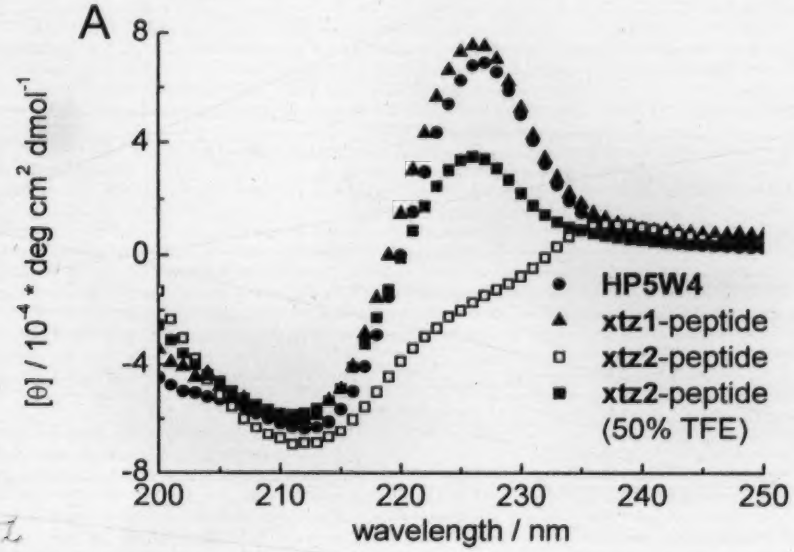
[a] Rates are the initial velocities of trypsinolysis divided by the total number of lysine and arginine residues in the peptide sequence, relative to **tz1** which was assigned a value of 1.00. [b] Errors for three or more independent measurements are reported in parenthesis as ± standard deviation.

The *in vitro* FRET efficiencies and relative rate of trypsinolysis were determined and shown in **Table 7**. The experimentally determined FRET efficiency was highest for the **xtz1**, and decreased with more destabilizing substitutions presented. By analogy with the **tz** and **xtz2** series, which contained variants with destabilizing mutations [1, 2], we can conclude a stability ranking of (most stable) **xtz1**>**xtz2**>**tz1**. Indeed, even when five destabilized amino-acids substituted, **xtz1** still showed a higher FRET efficiency while compared to both **tz1** and **xtz2** with no destabilizing substitution. This great stability

makes **xtz1** peptide a particular good peptide scaffold that can tolerate non-Trp-face residues randomization.

4.3.3 Characterization of the isolated peptide portion

The isolated peptide portions of **tz1** (**HP5W4**), **xtz1** (**xtz1**-peptide), and **xtz2** (**xtz2**-peptide) were produced either through peptide synthesis or by recombinant expression (Section 2.2.3). Similar to previously reported for highly-folded trpzip peptides [146], both **HP5W4** and **xtz1**-peptide are soluble in water at millimolar concentrations and have characteristic exciton-coupled bands in their CD spectra (**Figure 28A**). In contrast, **xtz2**-peptide is soluble in water at only micromolar concentrations and in the absence of TFE, its CD spectrum showed no significant exciton peaks [2]. We attribute the lack of solubility in water to the hydrophobic character conferred by the additional Trp side chains present in **xtz2**-peptide. It is apparent that under conditions more favourable to its dissolution, such as 50 % TFE or when CFP and YFP are fused to the termini, the peptide can indeed fold into a trpzip-type structure.



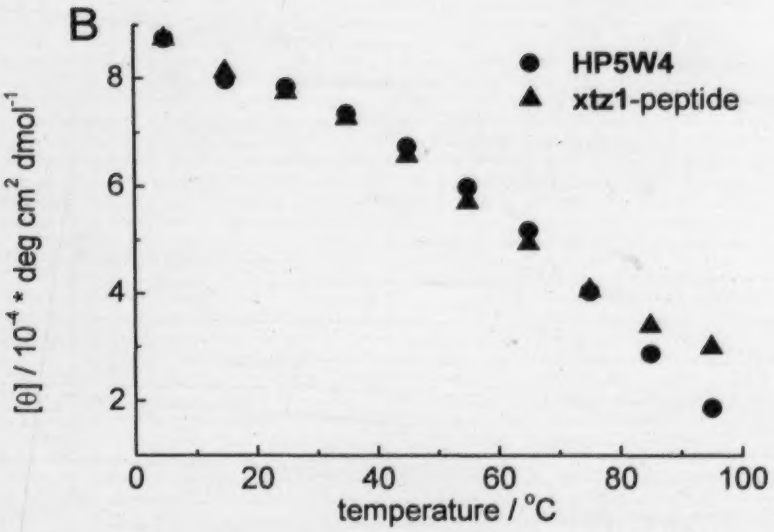


Figure 28 CD spectra and melting curves of **HP5W4**, **xtz1**-peptide and **xtz2**-peptide. (**A**) CD spectra of identical concentrations of **HP5W4**, **xtz1**-peptide, **xtz2**-peptide in buffer and **xtz2**-peptide in 50% TFE. The strong peak at 227 nm is a characteristic feature of trpzip-type peptides and is attributed to exciton coupling of the closely packed tryptophan residues [146]. Only in the presence of TFE does **xtz2**-peptide exhibit a significant exciton peak. (**B**) CD melting curves at 227 nm for **HP5W4** and **xtz1**-peptide.

4.3.4 xtz1 structure elucidated by NMR

To obtain further insight into the structure of **xtz1**-peptide, 2D NMR spectra were acquired and NOE distance constraints were employed to generate a structure ensemble (**Figure 29**).

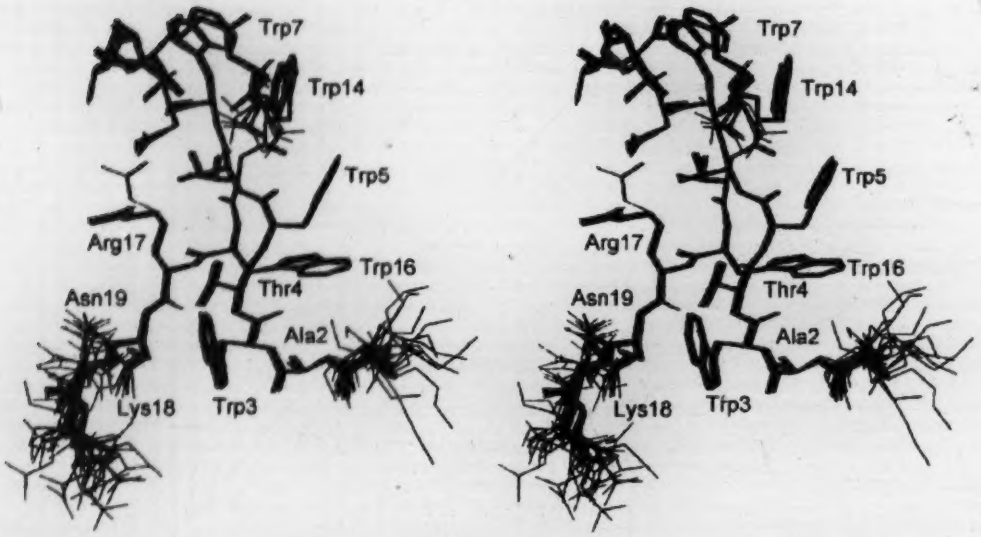


Figure 29 Stereoview of the NMR structure ensemble of **xtz1**-peptide with labels indicating the four tryptophan residues derived from **HP5W4** and the six residues identified at randomized positions. Coordinates have been deposited in the Protein Data Bank (ID: 2ORU)

As expected, **xtz1**-peptide adopts a typical trpzip-type β-hairpin conformation with the indole rings of the four **HP5W4**-derived Trp residues tightly packed on one face of the structure. The relative orientation of the diagonally paired Trp residues closest to the turn (Trp7/Trp14) is very similar to the edge-to-face (EtF) interaction observed in **trpzip4**, **HP5W4**, and HP7 (**Figure 30**) [146, 147]. The middle pair of indole rings (Trp5/Trp14) displays a parallel-displaced orientation similar to that observed in **HP5W4** [147]. It has been reported that the diagonal Trp pair closest to the termini (Trp5/Trp16) displays an EtF interaction in **trpzip4** but a parallel-displaced interaction in **HP5W4** [147]. In **xtz1**-peptide, Trp5/Trp16 are oriented in what is best described as an EtF interaction. This adjustment in packing interaction (relative to **HP5W4**) is attributed to the additional conformational restrictions enforced by the additional cross-strand interactions associated with the inserted residues.

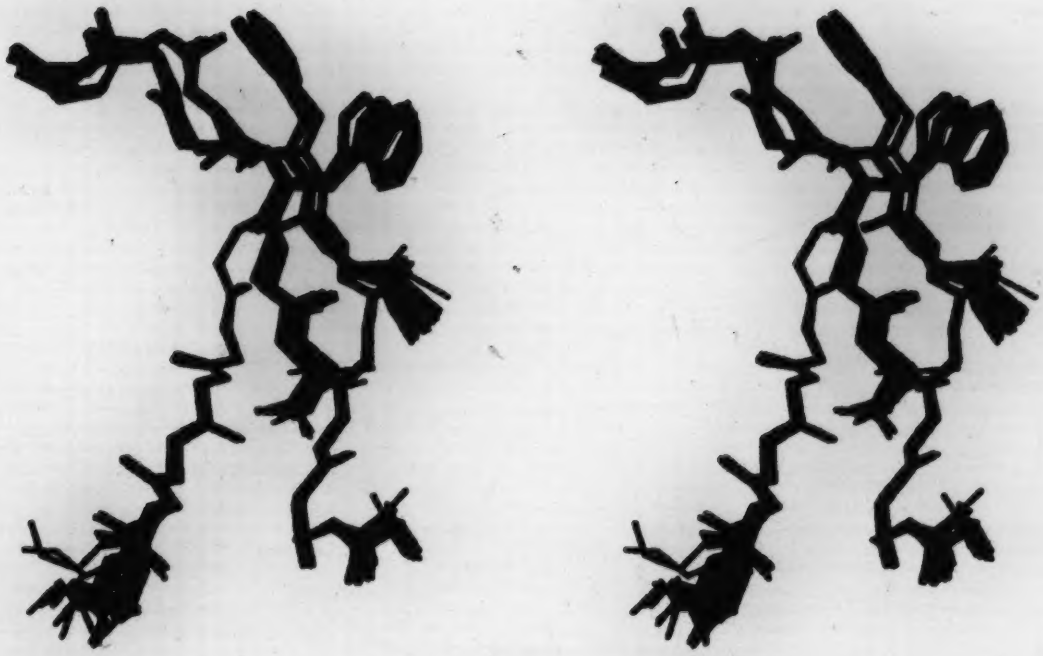


Figure 30 Stereoview of a superposition of the NMR structure ensembles of **HP7** (red) and **xtz1**-peptide (blue). All atoms from Thr2 to Thr11 of **HP7** (PDB: 2EVQ) [147] were aligned with all atoms from Thr6 to Thr15 of **xtz1**-peptide using the least squares fit function of XtalView v.4.0 [186].

During the three rounds of peptide-library screening, the ratio of YFP-to-CFP fluorescence emission intensity for the most abundant variants increased from 7.0 to 11.0 to 11.5. The structure of **xtz1**-peptide provides insight into the specific molecular interactions that are associated with these increases in FRET efficiency. For example, in round 1, we found that the peptide sequences that gave rise to the highest FRET efficiency had Trp at position 3 and Lys or Arg at the cross-strand diagonal position 18. This preference implied the presence of a Lys3/Trp18 cation-π interaction in **xtz1**-peptide [187-191]. Indeed, analysis of ^{1}H chemical-shift deviations in **xtz1**-peptide revealed that the side chain of Lys18 is significantly more shielded than any other non-Trp residue in the peptide (**Figure 31**). This result is consistent with the observed packing of the Lys18

side chain against the indole of Trp3 in a canonical cation-π interaction (**Figure 32A**) [187, 188].

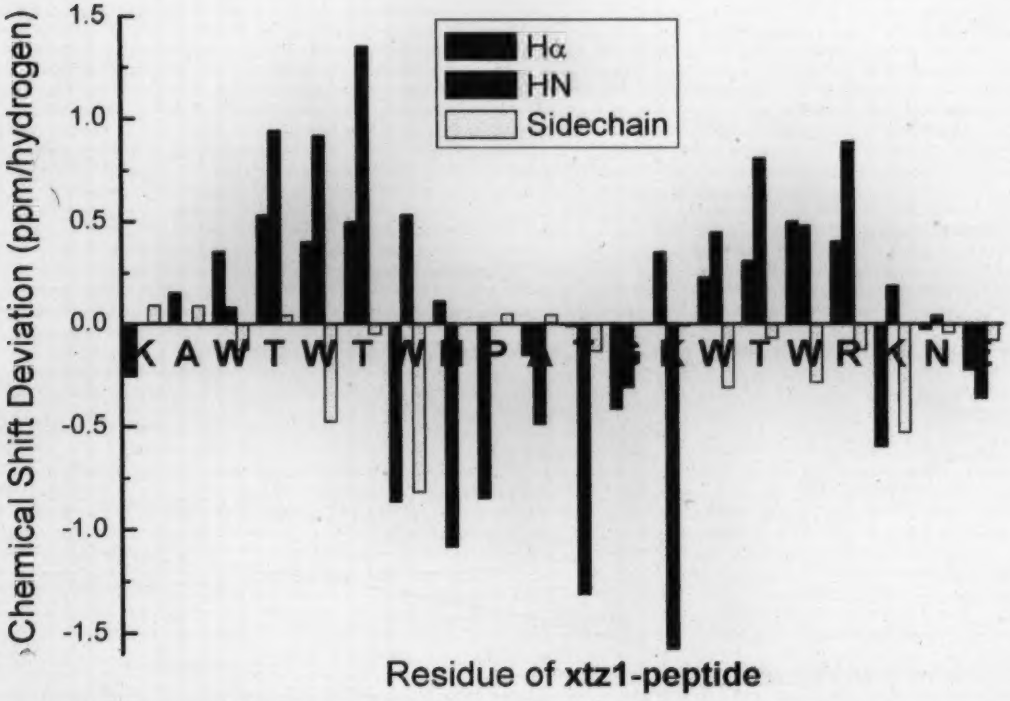


Figure 31 Chemical shift deviation (CSD) analysis of Hα, HN, and side chain hydrogen atoms of **xtz1**-peptide. CSDs were calculated by subtracting the residue-specific random coil chemical shift [192] from the observed chemical shift. For the side chain values, CSDs were summed over all hydrogen atoms in the residue and then divided by the total number of hydrogen atoms. The sign (positive = downfield shift; negative = upfield shift) and magnitude of the CSDs are in excellent agreement with the values previously reported for similar peptides [147, 155]. Note the strong upfield shift of 0.5 ppm per hydrogen over the complete side chain of K18.

In round 2, there was a strong preference for small residues (such as Ala and Ser) at position 2 and Asn at position 19. Rationalization of the position 19 preference is difficult, since the side chain is disordered in the NMR ensemble (**Figure 29**). For position 2, the structure suggests that the preference could be attributed to steric

hindrance between the larger side chains and the indole rings of Trp3 and Trp16 (**Figure 29**). In round 3, a strong preference for Thr or Ser at position 4 and Arg or Lys at position 17 was observed. It is speculated that only Thr or Ser at position 4 would be able to participate in an apparent cross-strand hydrogen bond between the side-chain hydroxyl and the main-chain carbonyl of Arg17 (3.5 Å, **Figure 32B**). The side chain of Arg17 extends along the groove between the two strands of the β-hairpin, making hydrophobic contacts with Thr4 and Thr6 and hydrogen bonds to the main-chain carbonyl of Trp7 (3.1 Å, **Figure 32B**) and the side-chain hydroxyl of Thr15 (3.4 Å, **Figure 32B**). A lysine at this position could presumably make similar contacts.

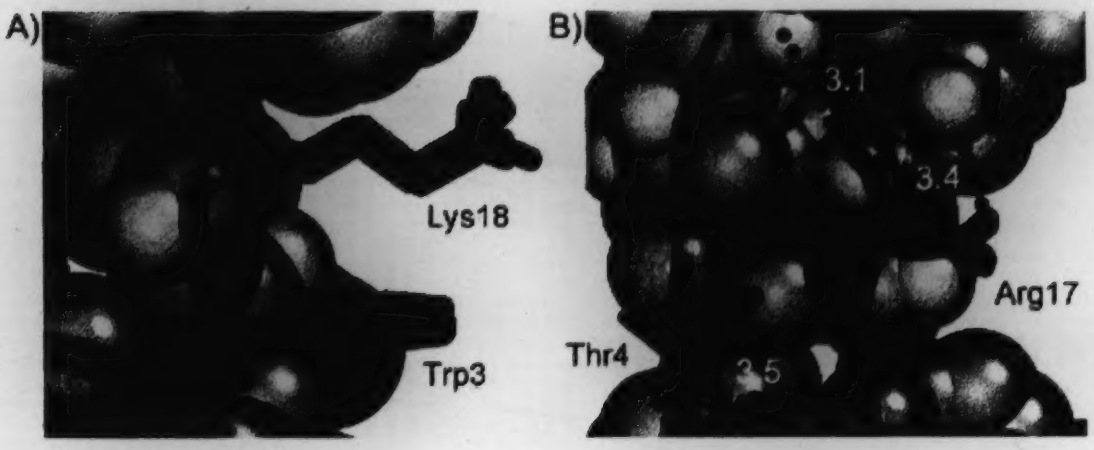


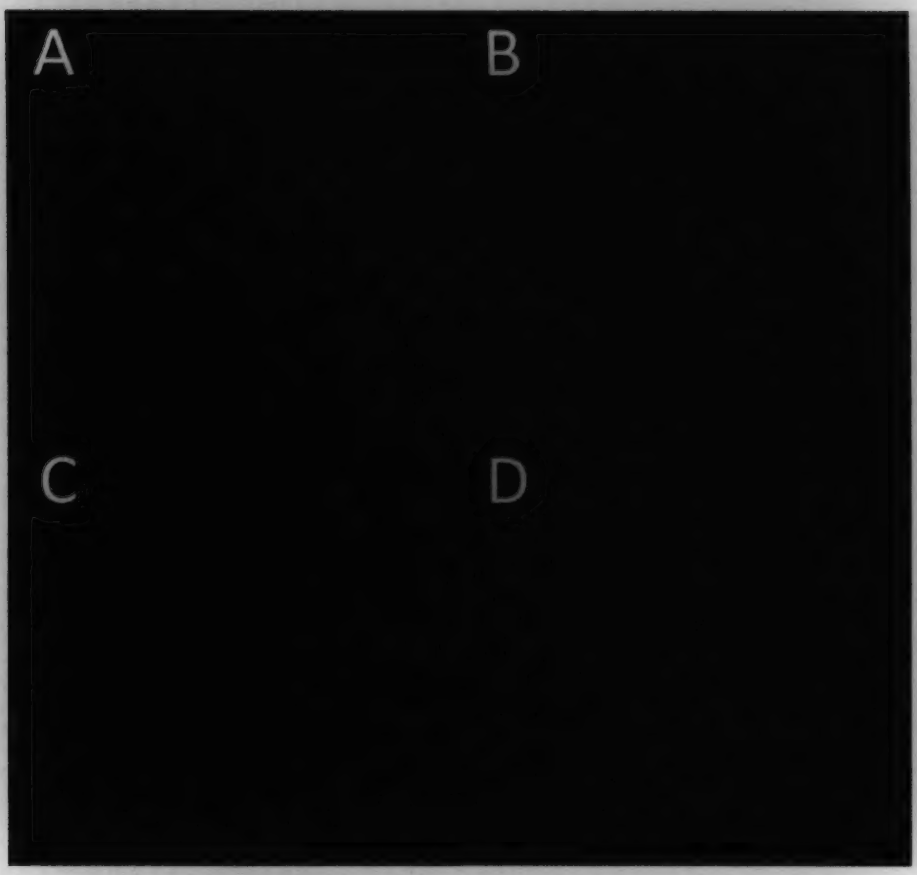
Figure 32 A close-up look at the "extended" residues. (**A**) Detail of the Trp3/Lys18 cation-π interaction in all 20 structures of the ensemble. (**B**) Detail of the Thr4/Arg17 pair with important hydrogen-bond interactions indicated with a dashed line. Typical O-O or O-N distances [Å] are shown in white. The remainder of the peptide is shown as a single representative space-filling structure.

We surmise that the combination of residues selected at the targeted positions of **xtz1**-peptide is particularly effective at holding the ends of the peptide in close proximity and minimizing the average distance between the fused CFP and YFP so that they exhibit more efficient FRET. In other words, these residues are participating in specific

molecular interactions that stabilize a well-folded conformation of the "extension" that we have appended onto the parent hairpin, **HP5W4**. Why then does this additional stabilization not manifest itself as a higher melting temperature for **xtz1**-peptide relative to **HP5W4** (**Figure 28B**)? We speculate that the answer lies with the fact that the fitness criterion used during library screening was high FRET efficiency, not overall thermal stability of the hairpin fold. Consequently, if the fraction of peptide folded at ambient temperature is not significantly diminished, a partially destabilized hairpin could still be "improved" by the FRET fitness criterion. We expect that this situation is occurring in **xtz1**-peptide; overall fold stability has been partially sacrificed in order to maximize FRET efficiency. Specifically, the Lys3/Trp18 cation-π interaction has introduced conformational restrictions that force the indole side chains of Trp5 and Trp14 to pack in a conformation that is of higher energy than the conformation they adopt in **HP5W4**. It has previously been suggested that there are intrinsic limits on β-strand length for some sequences [193]. Since the extension present in **xtz1**-peptide is not in a β-sheet conformation, our results neither support nor controvert this suggestion.

4.3.5 Assessing the conformational stability in mammalian cells

Having established that **xtz1**-peptide is highly folded in *E. coli* and *in vitro*, we transfected HeLa cells with plasmid DNA encoding **tz1**, **xtz1**, **xtz2**, and **xtz3** to determine if their conformational stability would be retained in mammalian cells. Fluorescence imaging revealed that **tz1**, **xtz1** and **xtz3** are evenly distributed throughout the cell (**Figure 33A, 33B and 33C**) and exhibit FRET efficiencies that closely parallelour *in vitro* results (**Figure 33E**). In contrast, much of the **xtz2** protein formed a low FRET efficiency (comparable to the unstructured peptide **xtz3**) aggregate in the apparent vicinity of the endoplasmic reticulum (**Figure 33D**). The low FRET efficiency of the aggregated protein might be due to either a lack of peptide structure in the fusion protein or a loss of FRET due to proteolysis. The **xtz2** excluded from the aggregate had a FRET efficiency similar to the *in vitro* value.


A
B
C
D

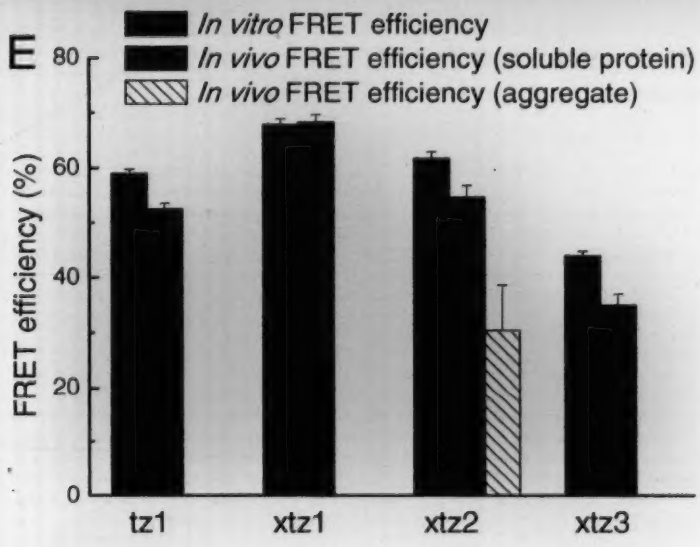

E
In vitro FRET efficiency
In vivo FRET efficiency (soluble protein)
In vivo FRET efficiency (aggregate)
FRET efficiency (%)
0
20
40
60
80
tz1
xtz1
xtz2
xtz3

Figure 33 Fluorescence imaging of the **tz1**, **xtz1**, **xtz2**, **xtz3** FRET constructs and the FRET efficiencies comparisons *in vitro* and in HeLa cells. Fluorescence imaging of the FRET constructs (**A**) **tz1**; (**B**) **xtz1**; (**C**) **xtz2**; and (**D**) **xtz3** expressed in HeLa cells. Each panel is a composite of the CFP fluorescence and a differential interference contrast image. While the fluorescence of **tz1**, **xtz1** and **xtz4** is distributed evenly throughout the cell, that of **xtz2** appears to aggregate in the vicinity of the nuclear membrane. (**E**) FRET efficiencies determined for proteins *in vitro* and in HeLa cells.

4.3.6 Beyond the 20-mer: efforts directed toward the development of 24-mer and 28-mer β-hairpins

The poor solubility of **xtz2**-peptide rendered it a "dead-end" for development as a generic peptide scaffold or as the basis for further extension of the hairpin. In contrast, the peptide portion of **xtz1** does appear to hold a lot of promise for further work in these directions. For these reasons, the peptide portion of **xtz1** has served as the progenitor for our continuing efforts to create highly folded 24-mer and 28-mer β-hairpin peptides.

In our first such effort, we constructed a library of protein variants (**Library 1, Table 8**) in which 2 partially-randomized residues were inserted before the 3rd and after the 18th residues of the peptide portion of **xtz1**. The particular subsets of amino acids represented at each position were chosen because these amino acids had frequently been identified at cross-strand positions during screens directed towards the development of **xtz1**. Several plates of colonies expressing members of **Library 1** were imaged and the YFP/CFP ratios determined for each individual colony. Disappointingly, the highest "in colony" ratios observed were ~3; only slightly greater than the "in colony" ratio of ~2 observed for similar constructs (such as **xtz3**) that have unstructured peptides linking CFP and YFP.

Table 8 Sequences of the 24-mer and 28-mer peptide libraries

Name	Peptide sequence fused between CFP and YFP [a]	YFP /CFP ratio	Comments
Library 1	KA--**UZ**WTWTWNPATGKWTWRK**ZU**--NE	≤ 3[c]	Further extension of **xtz1**; no clones picked.
Library 2	KA--**X**TWTWTWNPATGKWTWRKT**X**--NE	≤ 3[c]	Further extension of **xtz1**; the two highest ratio clones (**L2a**) were identical.
Library 3	KA--**XX**WTWTWNPATGKWTWRK**XX**--NE	≤ 5[c]	Further extension of of **xtz1**; clones **L3a-e.**
Library 4	KKWTWT**X**TWTWNPATGKWTWT**X**TWTWQE	≤ 2[c]	Clones picked but no soluble proteins were identified.
L2a	KA--WTWTWTWNPATGKWTWRK**TL**--NE	2.6[b]	Soluble.
L3a	KA--**N**PWTWTWNPATGKWTWRKNF--NE	4.0[b]	Soluble.
L3b	KA--**SP**WTWTWNPATGKWTWRK**NS**--NE	4.5[b]	Soluble.
L3c	KA--**NP**WTWTWNPATGKWTWRK**TH**--NE	4.8[b]	Soluble.
L3d	KA--**RD**WTWTWNPATGKWTWRK**ES**--NE	5.2[b]	Soluble.
L3e	KA--**NP**WTWTWNPATGKWTWRK**EG**--NE	5.5[b]	Soluble.

[a] **U** = amino acids Trp/Lys/Arg/stop, (codon WRG); **Z** = amino acids Ala/Val/Gly/Thr/Met/Arg (codon "RBG"); **X** = all 20 amino acids (codon "NNK"). Underlined residues are mutations or insertions relative to **tz1**. Bolded amino acids were selected from library screening. [**b**] Ratios are determined with soluble protein extracted from overnight cultures of single clones of bacteria. [**c**] Ratios are "in colony" measurements and are a reliable measure of relative FRET efficiency when compared to other "in colony" measurements.

For our second attempt, we constructed a similar library of protein variants (**Library 2**, **Table 8**) in which the inserted residues directed towards the non-Trp-face were threonine and the inserted residues directed towards the Trp-face were completely randomized. As with **Library 1** the highest "in colony" ratios were only ~3, however we did pick and

sequence two of the highest ratio clones. Interestingly, both clones (**L2a**, **Table 8**) were found to have identical residues at the randomized positions; a tryptophan at position 3 and a leucine at position 22. The hydrophobic character of both residues, and their expected close proximity in cross-strand positions, suggests that these two residues could be stabilizing the hairpin structure through Van der Waals' contact.

For our third attempt, all four inserted residues were randomized to all possible amino acids (**Library 3**, **Table 8**). We were encouraged to observe a range of "in colony" YFP/CFP ratios up to values of ~5 for this larger library. Five clones (**L3a-e**, **Table 8**) with high YFP/CFP ratio were picked and the DNA sequenced to reveal the composition of the peptide sequence. The peptide sequences did not reveal a single clear consensus, though there was a strong preference for asparagine and proline at positions 3 and 4, respectively. However, none of these 24mer clones exhibited a YFP/CFP ratio as high as that observed for **xtz1**.

In our sole attempt at creating a highly folded 28mer hairpin sequence, we semi-rationally designed a library of trpzip-type variants with 4 Trp/Trp cross-strand pairs interrupted by a central cross-strand Trp-face pair of randomized residues (**Library 4**, **Table 8**). The "in colony" ratios for this library were very low (~2). Several clones were picked but it was not possible to extract any soluble protein from overnight cultures.

4.4 CONCLUSION

This conclusive demonstration of the efficacy of *in vivo* screening provides researchers with a powerful new tool in the search for folded proteins [156, 172, 173]. The 20-mer **xtz1** peptide identified from the larger libraries showed promises as a generic protein scaffold as it can tolerate destabilizing substitutions while retaining its high solubility. A peptide library was created by randomizing six non-Trp-face residues (discussed in **Chapter 5**). This peptide library was subsequently screened for molecular recognition properties.

We further attempted to extend this 20-mer peptide to be 24 or 28-mer long, but no promising candidates were found. With the benefit of hindsight and experience, we expect that the most effective approach for development of 24-mer or longer hairpin peptides will be an incremental approach with extensive optimization of the intermediates. Therefore, this combinational library screening approach will be carried forward to identify hairpin peptides with increased stability (discussed in **Chapter 5**).

A direct application of this high-throughput *in vivo* screening method is to screen large number of variants to identify more extended and folded β-hairpins. Although my efforts were focused on tryptophan zipper β-hairpin peptides, I expect that this platform technology could be used to screen libraries of practically any small genetically-encoded peptide motif for variants that exhibit higher folding efficiency or stability in the cytoplasm of live cells.

CHAPTER 5:

ENGINEERED TRYPTOPHAN ZIPPER-TYPE PEPTIDES AS MOLECULAR RECOGNITION SCAFFOLDS

5.1 INTRODUCTION

Binding proteins and peptides that are capable of specific molecular recognition in the reducing environment of a cell's interior are valued as research reagents or potential therapeutic leads [109, 134, 135]. Antibodies are, ostensibly, the protein family of choice for most biological molecular recognition applications. These versatile molecular recognition reagents can be readily generated either by traditional immunization or hybridoma techniques [66] and provide researchers with high specificity and high affinity research tools. However, their use in live cell intracellular applications is generally limited by the technical challenge of getting polyclonal or monoclonal preparation of these large proteins across the cell membrane. For research applications (as opposed to therapeutic applications) this problem could, in principle, be circumvented by recombinant expression of genes encoding the heavy and light chains of the antibody in the target cells. However, this approach is limited by the poor folding efficiency of antibodies in the reducing environment of the cytoplasm due to an inability to form the multiple disulfide bonds that are critical for their stability [139]. Much work has been invested in the development of recombinant engineered antibody fragments [138] that can be expressed in living cells and retain the binding specificity and affinity of the intact antibody from which they were derived [194, 195]. An alternative approach is to engineer non-immunoglobin proteins or peptides to have molecular recognition functions that are not dependent on the formation of disulfide bonds [74, 140].

As discussed in **Chapter 1**, non-immunoglobulin domain proteins that are best suited for conversion into molecular recognition domains have been referred to as "generic protein scaffolds" [74, 140]. Desirable features of a generic protein scaffold intended for intracellular applications include a relatively small size and the absence of a disulfide bond (**Chapter 1**). An intriguing minimal scaffold for the intracellular display of constrained peptides is a "stem-loop" type structure [141, 142]. Such a structure has been previously proposed for an engineered peptide composed of two copies of an antiparallel dimerizing peptide (sequence SKVILFE) derived from Neuropeptide Head Activator [196], which presumably forms a 2-stranded β-sheet (the stem), flanking an unstructured

sequence of randomized residues (the loop) (**Figure 34**) [142]. Although this design is compelling, evidence supporting the existence of the stem-loop peptide structure is inconclusive. Indeed, an NMR study by Lai and Gellman [197] led the authors to conclude that, even with a strongly β-hairpin-promoting D-Pro-Gly sequence in place of the unstructured loop, the SKVILFE stem structure is only moderately stable. A recent CD study by Anderson and coworkers [141] led the authors to conclude that tandem copies of EFLIVKS (the retro sequence of Neuropeptide Head Activator) can serve to tether together the ends of an intervening 18 residue-unstructured loop.

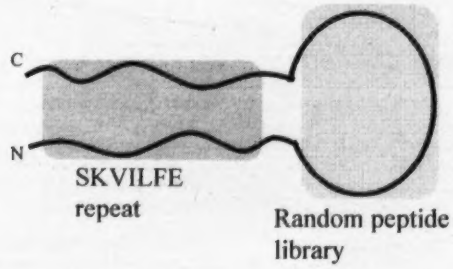


Figure 34 Schematic representation of "stem-loop" type structure based on SKVILFE repeat.

It had occurred to us that, regardless of whether the Neuropeptide Head Activator or its retro sequence is able to self dimerize, there may exist other peptide sequences that are more effective at forming the stem portion of the stem-loop peptide structure. As described in previous chapters, trpzip is a very promising candidate for this particular application. Its remarkable folding stability arises from the $\pi - \pi$ interaction in the two interdigitating cross-strand Trp-Trp pairs [146]. We reasoned that it might be possible to find a particularly stable variant that would retain the β-sheet structure even upon replacement of the optimized turn with a larger and unstructured loop. Thus, this variant could form the "stem" scaffold for randomized peptides presentation and selection for target binding by phage display [69]. Alternatively, a structured trpzip peptide could also

be a scaffold for multiple randomized amino acids presentation at the non-Trp-face residues (described in **Chapter 4**).

We had developed an *in vivo* method for rapid high-throughput screening of thousands of peptide sequences to find ones that are highly structured in live cells [1, 2]. The detailed screening strategies were discussed in **Chapter 2** and **Chapter 4**. We had applied this screening method to the **HP5W4** peptide in an effort to engineer additional stabilizing interactions into this already-very-stable hairpin. Our strategy involved extending the sequence by two randomized residues in each strand of the β-hairpin. Iterative library screening resulted in the identification of the 20-mer **xtz1**-peptide (**Chapter 4** and **Table 9**) with an additional cross-strand lysine-tryptophan cation-π interaction relative to **HP5W4** [2]. In this chapter, I report our continued efforts to engineer additional stabilizing interactions into the **xtz1**-peptide through the use of our *in vivo* screening approach. The latter sequence was extended to 24 residues with exhaustive screening of all possible amino acid combinations at each cross-strand pair of residues. Using our most promising 24-mer peptide, we replaced the central two residues of the optimized turn sequence with seven glycine residues as an unstructured "loop". Results obtained from our FRET-based assay support the conclusion that this sequence is able to adopt a highly folded structure that we conceptualize as a stem-loop.

Peptide libraries were created based on the **xtz1** and **xxtz1** scaffolds (**Figure 35**), with randomized amino-acids substitution at the non-Trp-face-residues or randomized loop insertion, respectively. These minimal and disulfide bond-free scaffolds were subsequently screened for molecular recognition properties by phage display.

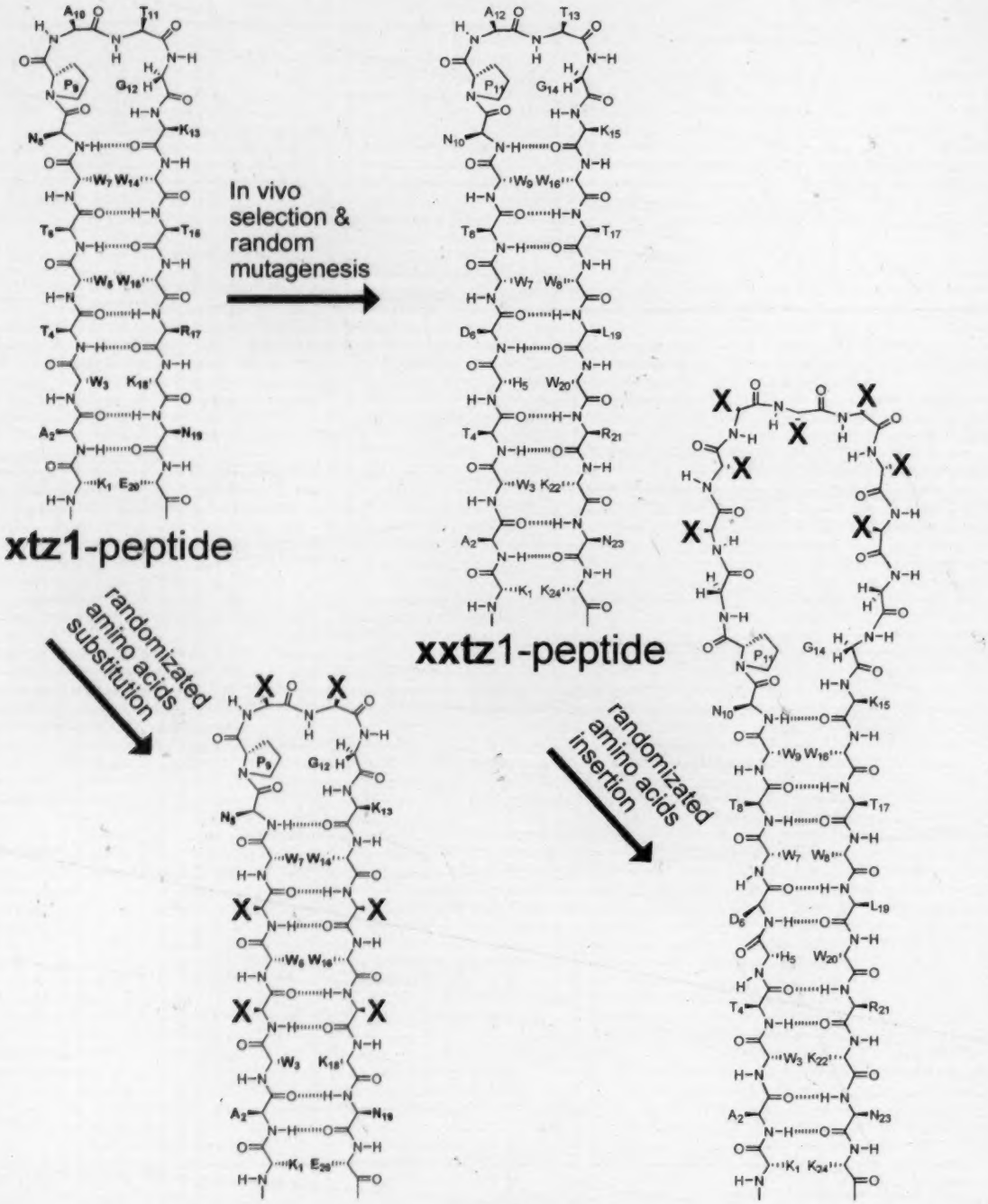


Figure 35 Schematic illustration of the **xxtz1**-peptide engineering and the phage display libraries used in **Chapter 5**. The **xtz1**-peptide based phage display library was

generated by substituting randomized amino acids on the non-Trp-face residues. The **xxtz1**-peptide based library was constructed by inserting a seven-member randomized loop and replacing the residues 12 and 13. The randomized residues are represented with a red "X".

5.2 MATERIAL AND METHODS

5.2.1 General procedures and materials

Please refer to **Section 2.2.1**.

5.2.2 Library construction and screening

The CFP-peptide-YFP libraries were expressed using the previously described vector pZC1 (**Section 2.2.2**). The 24-mer peptide libraries for screening were created either by PCR method or annealing two complementary single stranded oligonucleotides, methods that have been described in **Section 4.2.2**. Error prone PCR (PCR reactions with 0.15 mM Mn^{2+} ion present and with one dNTP at 0.025 mM and three dNTPs at 0.25 mM) were employed to increase library diversity in the last three rounds of screening. The dsDNA was ligated into the CFP-peptide-YFP expression vector that had been digested with *Sac*I and *Kpn*I restriction enzymes. The system for imaging the fluorescence of bacterial colonies grown on 10 cm Petri dishes is described in **Section 2.2.6**. Individual colonies that had the highest ratio of YFP to CFP emission and the highest brightness were picked from the plate and cultured. The sequence of the peptide portion of all selected proteins was determined by DNA sequencing with a forward primer (5'-CCCTCGTGACCACCCTGACCTGG-3') that anneals to the chromophore region of the gene for CFP. The primer used for sequencing of the displayed portion of phagemid pComb3ZC was 5'-AGGCTTTACACTTTATGCTTCCGGC-3'.

5.2.3 Protein and peptide purification

For protein purification, please refer to **Section 2.2.4**.

The **xxtz1**-peptide was isolated using a previously described bacterial expression system **Section 2.2.3**. The **xxtz1(loop)** peptide (**Table 10**) was synthesized on a ABI 433A peptide synthesizer (Applied Biosystems). The peptides were further purified by reversed-phase chromatography on a Prosphere HP C18 300A column (Alltech Associate, Inc.). A linear gradient of increasing acetonitrile in H_2O with 0.1% trifluoroacetic acid was used to elute the target peptides. The fraction containing the target peptide was lyophilized to provide a fluffy white powder. A calculated extinction coefficient of 33,450 $M^{-1}cm^{-1}$ at 280 nm was used to determine the concentration for both of the peptides.

5.2.4 Spectroscopy

Please refer to **Section 2.2.5**.

5.2.5 Phage displayed peptide library construction and screening

The plasmid vector used for cloning and phage display is referred to pComb3ZC. It was constructed as follows. The phagemid pComb3HTT [198] (kindly provided by Carlos F. Barbas, the Scripps Research Institute, La Jolla, Calif.) was digested with *Sfi*I to prepare the linear vector with sticky ends. The digested pComb3HTT vector was ligated with a synthetic dsDNA that encodes a linker region with multiple restriction sites and has appropriate sticky ends. The resulting vector (pComb3ZC) had a convenient set of restriction sites (*Sac*I/*Kpn*I/*Xba*I/*Xho*I/*Spe*I) for library insertion.

The **xxtz1**-peptide-based phage display library, which encodes 7 randomized amino acids in the "turn" region (**Table12**), was generated by a PCR approach. A ssDNA encoding

the desired randomized residues (5'-C TGG ACT TGG AAT CCC GCC $(NNK)^7$ GGT GGC AAG TGG ACA TGG C-3') were PCR-extended with a 5' primer (5'-GGT GAG CTC AAA GCT TGG ACC CAC GAC TGG ACT TGG AAT CCC GGC-3') and a 3' primer (5'-AAC CTC GAG TTT GTT CTT TCT CCA CAG CCA TGT CCA CTT GCC ACC-3'). The PCR product was purified using a QIAEX II Gel Extraction Kit (Qiagen) and digested with *Sac*I/*Xho*I. The digested product was ligated into the pComb3ZC vector which had been digested with the same two restriction enzymes.

The **xtz1**-peptide-based phage display library encodes six randomized amino acids on the non-Trp-face residues of the peptide (**Table 12**). The library was constructed by a PCR primer extension reaction with two primers hybridizing to each other via 17 base pairs (5'-GTT GAG CTC AAG GCC TGG NNK TGG NNK TGG AAT CCC NNK NNK GGC AAA TGG -3' and 5'- ACC CTC GAG CTC GTT CTT MNN CCA MNN CCA TTT GCC MNN MNN GGG ATT CC-3'). The ligated products were used to transform *E. coli* XL1-blue cells (Stratagene) by electroporation. Appropriate dilutions of the transformed bacteria were plated on LB/agar plates containing carbenicillin (0.1 mg/mL) and incubated at 37 °C overnight. The final **xxtz1**-peptide based library consisted of 4×10^8 clones, and the **xtz1**-peptide based library consisted of 7×10^7 clones.

Phage production and enrichment were performed using established protocols [124]. Briefly, phagemid DNA were transformed into *E. coli* XL1-blue cells and grown in super broth medium (SB) containing carbenicillin (50 μg/ml) and tetracycline (10 μg/ml) at 37 °C for 2 hours. The culture was then infected with the VCSM13 helper phage (Stratagene). Following an incubation period of 1.5 hours, kanamycin was added to a final concentration of 70 μg/ml to select for helper phage infected cells, and the culture was incubated overnight at 37 °C with shaking at 300 rpm. The bacteria were pelleted by centrifugation (15 min, 10,000 rcf). Phage particles were precipitated from the culture supernatant by adding PEG-8000 (4% w/v) and NaCl (3% w/v) and harvested by centrifugation (30 min, 15,000 rcf). The phage pellet was resuspended in PBS buffer and was ready for use in panning experiments. The enrichment of streptavidin binding phage was performed on streptavidin magnetic beads (Invitrogen). Total 20 μl of streptavidin

magnetic beads were first incubated with BSA (3% w/v) in PBS for 1 hour. Approximately 5×10^{11} phage were then added and gently mixed with the streptavidin magnetic beads for one hour. The beads were washed with 10× with 1 ml PBS buffer containing 0.5% Tween20 to remove unbound and weakly bound phage. The adherent phage were then eluted with 0.1 mM biotin in PBS and amplified for subsequent selection. After 3 or 4 rounds of enrichments, 5 colonies of each library were cultured. DNA sequencing was used to reveal the gene sequence that corresponded to the displayed peptide portion.

5.2.6 Surface plasmon resonance

Kinetic analysis of **xxtz1-HPQ1** and its control protein (**Table 13** and **Figure 35**) binding to streptavidin was performed on a Biacore X instrument (Biacore Inc.) at 25 °C. A streptavidin-coated-chip (sensor chip SA, Biacore Inc.) was used for the measurements. HEP-EP buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) was used as the running buffer at a flow rate of 10 µl/min. Total 30 µl of freshly purified proteins (50 nM – 500 nM) in HEP-EP buffer were injected for each measurement. Dissociation constants (K_d) were calculated by BIAevaluation 4.1 software (Biacore Inc.) by fitting data to the Langmuir 1:1 binding model.

5.3 RESULTS AND DISCUSSION

5.3.1 *In vivo* screening to identify a 24-mer hairpin peptide

The template for the peptide library is a 24-mer version of our **xtz1**. This 24-mer contains two additional pairs of residues, one random and one threonine, genetically inserted after the 4th and 18th residues of **xtz1**. The sequence of the resulting "peptide" portion of the peptide library is $K^1A^2W^3T^4\underline{X}^5\underline{T}^6W^7T^8W^9N^{10}P^{11}A^{12}T^{13}G^{14}K^{15}W^{16}T^{17}W^{18}\underline{T}^{19}\underline{X}^{20}R^{21}K^{22}N^{23}E^{24}$, where X

represents all 20 amino acids. The imaging system and screening procedures were described in **Section 2.2.6**. Briefly, ~1 × 10^4 colonies were subjected to screening for each round. Colonies exhibiting high YFP-to-CFP fluorescent emission ratio and brightness were cultured and the plasmid DNA was purified.

Table 9 Peptide sequences identified by library screening in **Chapter 5**

library[a]	Sequences identified			
	X^5			X^{20}
	W			R
	W			R
	H			W
1 $K^1A^2W^3T^4\underline{X}^5\underline{T}^6....\underline{T}^{19}\underline{X}^{20}R^{21}K^{22}$	M			P
	K			P
	E			T
	G			P
2 $K^1A^2W^3T^4\underline{W/K/H/R/Y/G}^5\underline{X}^6....\underline{X}^{19}$	W/K/H/R/Y/G[5]	X^6	X^{19}	W/T/S/P/R[20]
$\underline{W/T/S/P/R}^{20}R^{21}K^{22}N^{23}E^{24}$	H	D	L	W[b]
	W	H	T	R
	W	D	L	W
	W	D	Q	W
	H	H	L	W
	K	G	V	W
	W	H	Q	W
3 Error prone PCR on **xxtz0.5**	$K^1A^2W^3T^4\underline{H}^5\underline{D}^6....\underline{L}^{19}\underline{W}^{20}R^{21}K^{22}N^{23}K^{24}\times2$ [c]			
4 Error prone PCR on **xxtz1**	no better variants were found			

[**a**] "..." denotes residues 7 to 18 of the peptide. Positions marked "X" were subject to saturation mutagenesis by using the codon "NNK" where N=adenine (A), guanine (G), cytosine (C), or thymine (T), and K=G or T. In the library 2, position 5 was mutated to Trp/Lys/His/Arg/Phe/Gly while position 20 was mutated to Trp/Thr/Ser/Pro/Arg by using different PCR primers. [**b**] This sequence was designated as **xxtz0.5**. [**c**] This sequence was designated as **xxtz1**.

Analysis of the DNA sequences revealed an apparent preference for aromatic and cationic pairs of residues at the cross-strand positions 5 and 20. The combined sequences from library 1 were used as the new template for the second library, in which positions 6 and 19 were randomized to any amino acid. Colonies were screened for high YFP-to-CFP emission ratio and the DNA of those colonies that exhibited the highest ratio was sequenced. Analysis of the DNA sequences revealed some strong amino acid preferences: His/Trp for position 5, Asp/His for position 6, Leu/Val/Thr for position 19, and Trp for position 20. By comparing FRET data from purified proteins (data not shown), we concluded that the sequence containing $H^5D^6...L^{19}W^{20}$, which was designated as **xxtz0.5** (**Table 9** and **10**), was the best candidate to be used as a template for future optimization by random mutagenesis. After 3 rounds of error-prone PCR, the optimized peptide **xxtz1** was obtained (**Table 9** and **10**).

5.3.2 Characterization of xxtz1 series proteins

Table 10 Summary of proteins and peptides compared in **Chapter 5**

Protein name	peptide name	peptide sequence (flanked by CFP and YFP in protein)
tz1	**HP5W4**	KKWTWNPATGKWTWQE
xtz1	**xtz1**-peptide	KAWTWTWNPATGKWTWRKNE
xxtz0.5	n/a	KAWTHDWTWNPATGKWTWLWRKNE
xxtz1	**xxtz1**-peptide	KAWTHDWTWNPATGKWTWLWRKNK
tz1-loop	n/a	KKWTWNPGGGGGGGGKWTWQE
xtz1-loop	n/a	KAWTWTWNPGGGGGGGGKWTWRKNE
xxtz0.5-loop	n/a	KAWTHDWTWNPGGGGGGGGKWTWLWRKNE
xxtz1- loop	n/a	KAWTHDWTWNPGGGGGGGGKWTWLWRKNK
n/a	**xxtz1(loop)**-peptide	KAWTHDWTWNPGSSGSSGGKWTWLWRKNK

In **Chapter 2**, I had reported a FRET method in monitoring genetically fused peptide structures. A similar strategy was employed to investigate **xxtz0.5** and **xxtz1** peptide structures in the context of CFP and YFP. As a comparison, previously reported FRET efficiencies of **tz1** and **xtz1**, as well as their non-structured counterparts **tz6** and **xtz3** (**Chapter 2** and **4**) are shown in **Table 11**. The higher FRET efficiency for **xxtz0.5** and **xxtz1** provides a good qualitative indication they are at least as structured as the previous versions (**Table 11**).

Table 11 **tz1**, **xtz1** and **xxtz** series proteins FRET comparison

	FRET efficiency [%]	FRET efficiency [%] with glycine loop inserted in the turn region
tz1	59.0 (±0.7)%	51.3 (±0.5)%
tz6	47.2 (±1)%	not investigated
xtz1	67.9 (±1)%	65.3 (±0.6)%
xtz3	44.0 (±0.8)%	not investigated
xxtz0.5	70.1 (±1)%	68.4 (±0.6)%
xxtz1	84.5 (±0.7)%	74.6 (±0.2)%

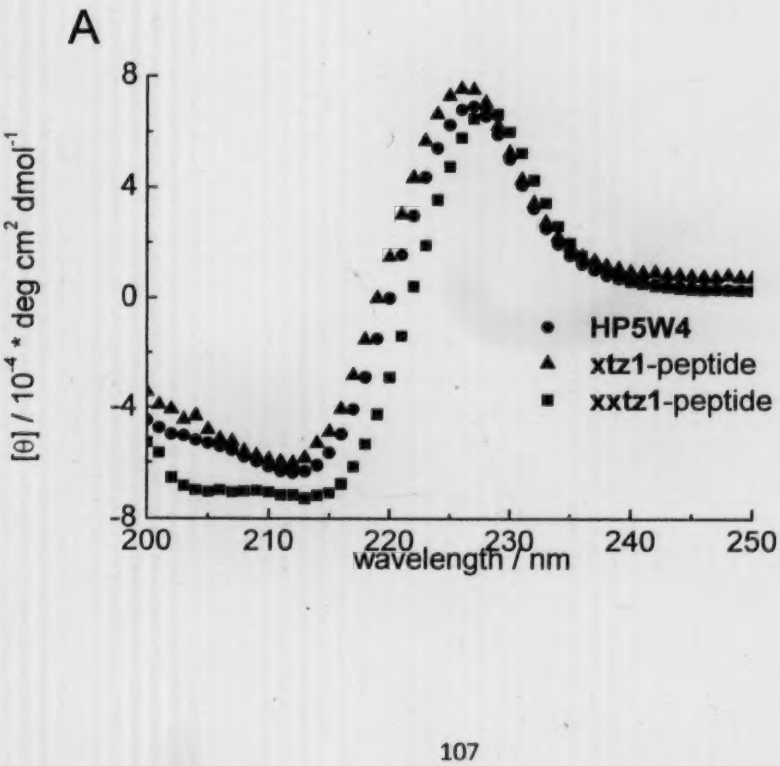
One must be careful in interpreting this result since there are at least 3 possible factors that may be contributing to the higher FRET efficiency. The first is that the folded structure of the peptide could bring the two randomly oriented FPs into closer proximity. The second is that the folding/unfolding equilibrium may lie more towards the folded structure. The third is that the structure of **xxtz1** rigidly biases the FPs to a relative orientation that is favorable to FRET (*i.e.*, $\kappa^2 > 2/3$). The important point is that all 3 factors necessitate that the peptide be highly folded in order to explain the higher FRET efficiency. Quantitative analysis of FRET efficiency is only valid across a conservatively and systematically varied series of peptides as described in **Chapters 2** and **4**.

One purpose of my research was to engineer a "stem" scaffold for randomized peptide presentation in order to select for high affinity target recognition. We had hypothesized that a highly structured trpzip-type variant would be a good "stem" scaffold candidate. So the next question we had to answer was whether our best variants could tolerate randomized amino-acids presentation. To test the variants' stability, a seven-glycine-residue "loop" was genetically inserted into **xxtz0.5** and **xxtz1** by replacing the previously optimized turn residues 12 and 13 (**Figure 35** and **Table 10**). These peptide sequences were genetically inserted between CFP and YFP in the pZC1 vector (**Section 2.2.2**). The FRET efficiencies of the resulting proteins, designated as **xxtz0.5-loop** and **xxtz1-loop** (**Table 10**), were measured (**Table 11**). As a comparison, this loop insertion strategy was also applied to **tz1** and **xtz1** (**Table 10 and 11**). When the loop was inserted, a decrease of FRET efficiency was observed for all of the proteins. However, none of the variants decreased as low as the unstructured control proteins, despite the fact that the overall inter-FP linker had increased by 7 residues in length. This result suggests that the peptide portion of these trpzip-type peptides remains highly structured in the context of two FPs even when the turn is replaced with an unstructured loop. Accordingly, these peptides are envisioned to adopt a stem-loop-type structure, indicating that they are promising candidates for grafting randomized peptide libraries into their turn regions. Due to the high FRET efficiency of the **xxtz1** with the glycine loop insertion, we choose to proceed with this variant. However, the other variants are equally promising as they are shorter and exhibited less of a decrease in FRET efficiency upon glycine loop insertion.

5.3.3 Characterization of xxtz1-peptide

To determine whether the **xxtz1** peptide was highly structured when removed from the context of the FP fusion protein, we expressed and purified the peptide alone (**xxtz1-peptide**). The isolated peptide was then analyzed by CD spectroscopy. Due to the indole side chains exciton coupling, the structure of trpzip peptides can be readily revealed by their unique CD signature absorbance at 229 nm [146]. The similar CD spectrum and the

signal magnitude of the **xxtz1**-peptide while comparing with **HP5W4** and **xtz1**-peptide (**Figure 36A**), confirms that the extra "extended" residues does not disturb the parental peptide structure. However, the CD melting curve revealed that the **xxtz1**-peptide has a melting point 15 °C lower than its ancestor **HP5W4** (**Figure 36B**). A small decrease in the melting temperature was also noticed for the **xtz1**-peptide, when compared to its precursor **HP5W4** (described in **Chapter 4**). As stated in the previous chapter, a FRET-based library screening is aiming for the peptides that have high FRET (which is presumably attributable to a close end-to-end distance or biased orientation), rather than the overall thermal stability. Based on these results, it seems that a peptide sequence with a high folding efficiency at ambient temperature does not necessarily have a higher melting temperature. As shown in **Figure 36B**, **HP5W4**, **xtz1**-peptide, and **xxtz1**-peptide all exhibit an effectively identical degree of melting between 5 °C and 20 °C. It is only at temperatures substantially higher than our screening temperature that differences in the peptide fold stability manifest themselves.



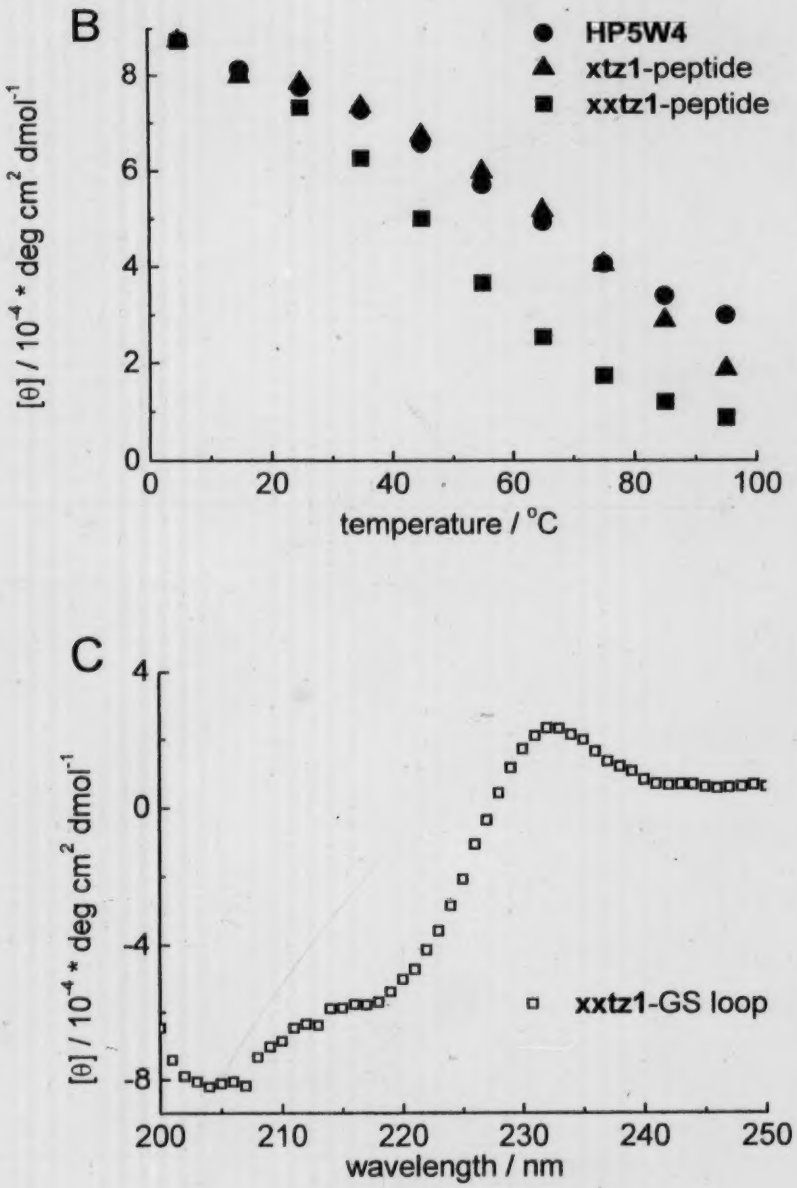


Figure 36 CD spectra and melting curves of **HP5W4**, **xtz1**-peptide, **xxtz1**-peptide and **xxtz1(loop)**-peptide. (**A**) CD spectra of identical concentrations of **HP5W4**, **xtz1**-peptide, and **xxtz1**-peptide. The strong peak at 227 nm is a characteristic feature of trpzip-type peptides and is attributed to exciton coupling of the closely packed tryptophan

residues [146]. (**B**) CD melting curves at 227 nm for **HP5W4**, **xtz1**-peptide and **xxtz1**-peptide. (**C**) CD spectrum of **xxtz1(loop)**-peptide.

To further investigate **xxtz1**-peptide stability in tolerating random loop insertion, we designed and synthesized a peptide with **xxtz1** framework where residues Ala12 and Thr13 were replaced with a glycine and serine rich seven-residue loop. This peptide was designated as **xxtz1(loop)**-peptide (**Table 10**). The **xxtz1(loop)**-peptide exhibits a much weaker exciton signature in its CD spectrum than the analogous peptide lacking the loop insertion (**xxtz1-peptide**, refer to **Figure 36A**). Furthermore, its signature CD absorbance is 5 nm red-shifted from that of **xxtz1**-peptide (**Figure 36C**). The trpzip characteristic CD signal at ~229 nm results from the Trp-Trp interactions [170]. The weak exciton signature of this CD spectrum indicates a substantial disruption of the Trp-Trp cross-strand packing and is consistent with the stem-loop peptide sequence being only partially structured. By taking the magnitude of the exciton in the CD as a rough indication of the concentration of folded protein, we estimate that the **xxtz1(loop)**-peptide is only 30% folded. This result does appear to be at odds with our FRET measurements. One explanation for the discrepancy is that the folding of the **xxtz1(loop)**-structure is highly context dependent. That is, in the context of the FP fusion protein the folding is improved due to improved solubility or possibly by molecular interactions with the FPs themselves.

5.3.4 Phage display libraries construction and screening

The goal of this work was to establish trpzip-type peptides as a scaffold for molecular recognition. As a proof-of-concept demonstration, **xtz1**-based and **xxtz1**-based libraries were constructed and displayed on the gIII protein of the M13 phage and panned with selection for streptavidin recognition. The **xtz1**-based library was constructed by genetically introduced six randomized amino acids at the non-Trp-face residues 4, 6, 10, 11, 15, and 17 (**Table 12** and **Figure 35**), while the **xxtz1**-based library was generated by

introducing a peptide loop with seven consecutive randomized amino acids and replacing the residues 12 and 13 in **xxtz1** (**Table 12** and **Figure 35**).

Table 12 Phage display libraries

	libraries[a]
xtz1-based scaffold	$K^1A^2W^3\underline{X}^4W^5\underline{X}^6W^7N^8P^9\underline{X}^{10}\underline{X}^{11}G^{12}K^{13}W^{14}\underline{X}^{15}W^{16}\underline{X}^{17}K^{18}N^{19}E^{20}$
xxtz1-based scaffold	$K^1A^2W^3T^4H^5D^6W^7T^8W^9N^{10}P^{11}G\underline{XXXXXXXX}GG^{14}K^{15}W^{16}T^{17}W^{18}L^{19}W^{20}R^{21}K^{22}N^{23}K^{24}$

[a] Positions marked "X" (underlined) were subject to saturation mutagenesis by using the codon "NNK".

A total number of 7×10^7 and 4×10^8 independent clones were obtained after the **xtz1** and **xxtz1** library transformations, respectively. It should be noticed that since the ideal **xtz1**-based library encodes $32^6 = 3.4 \times 10^{10}$ nucleotide sequences (the degenerate codon "NNK" has 32 combinations, but only encodes 20 amino acids), and the **xxtz1**-based library encodes 1.1×10^9 variants, the phage libraries we obtained did not contain all the possible peptide variants (the "obtained" library should be at least three times larger than the theoretical library in order for every possible variant presents). The phage library was then used to pan against streptavidin-immobilized magnetic beads. Streptavidin-immobilized magnetic beads, previously incubated with biotin, were used for panning as controls.

As shown in **Figure 37**, significant phage output increases were observed in the third round of selection for both libaries, indicating specific target binding phage enrichment [124]. This was further confirmed by comparison with the control experiments, in which no enrichment was noticed.

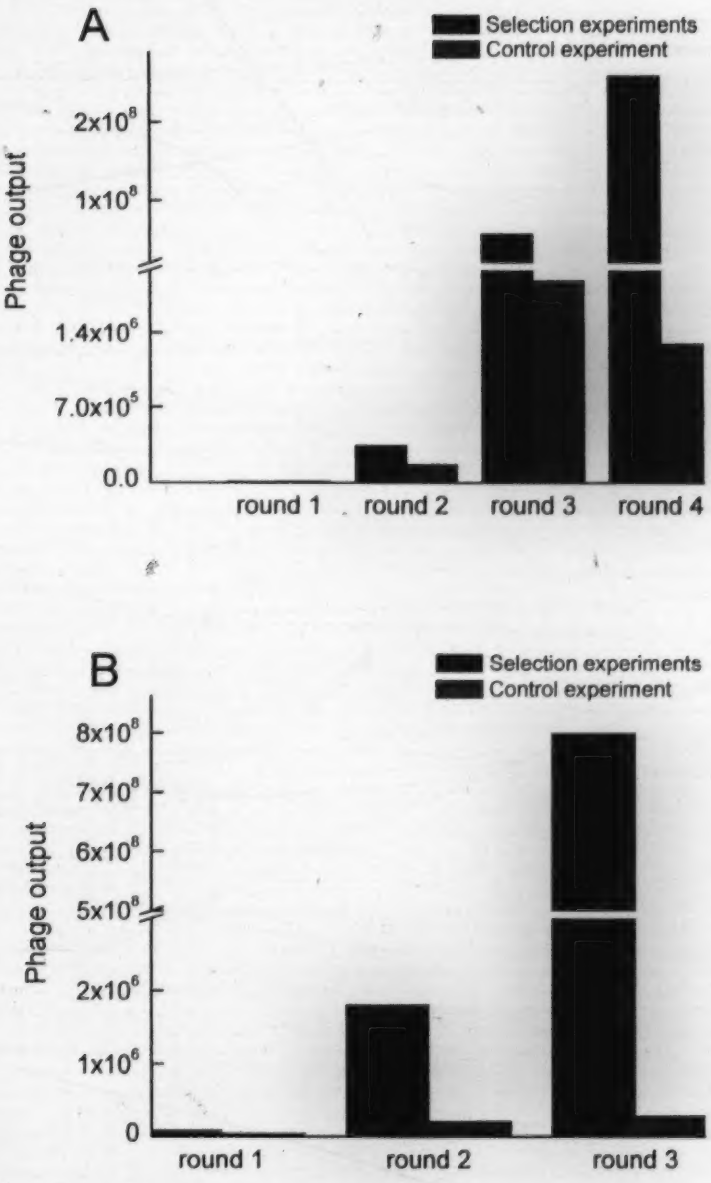


Figure 37 Phage output in each round of selection. **(A) xtz1**-based phage library; **(B) xxtz1**-based phage library. Phage experiment was carried out on streptavidin magnetic beads. The control experiments are identical to the selection experiment except pre-biotin-blocked magnetic beads were used.

It is well-known that streptavidin binding peptides typically share the common His-Pro-Gln (HPQ) motif [199]. This consecutive sequence theoretically should not be present in the **xtz1**-based library because of the restriction of randomized amino acid positions. After 4 rounds of selection, 5 clones were picked from the enriched **xtz1**-based library and the plasmids were sequenced. It is interesting to note that all five colonies had the same peptide sequence (**Table 13**), in which the HPQ motif is present! DNA sequencing revealed that the codon encoding Asn (AAT) at position 8 has a single base pair mutation (A to C) resulting in the codon CAT that encodes a histidine! This substitution might result from imperfect DNA nucleotide synthesis or error from PCR reaction. The normal Pro at position 9 and Gln at the randomized position 10 followed this replacement.

Table 13 Peptide sequences obtained from panning against streptavidin

Library	peptide sequence [a]	# of occurrence
xtz1-based library	KAW**P**W**Q**W**H**<u>**PQ**</u>SGKW**F**W**N**KNEGT	5
xxtz1-based library	KAWTHDWTWNPG**EMQDH**<u>**PQ**</u>GGKWTWLWRKNK[b]	2
	KAWTHDWTWNPG**EMDH**<u>**PQ**</u>NGGKWTWLWRKNK	1
	KAWTHDWTWNPG**ELDNH**<u>**PQ**</u>GGKWTWLWRKNK	1
	KAWTHDWTWNPG**EMSDH**<u>**PQ**</u>GGKWTWLWRKNK	1

[**a**] Randomized residue positions are in boldface type, and the streptavidin binding motif HPQ is underlined. [**b**] This peptide sequence was later genetically fused between CFP and YFP. The resulted protein was designated as **xxtz1-HPQ1**, the affinity of which to streptavidin was measured by surface plasmon resonance (SPR).

For the **xxtz1**-based library screening, 5 colonies were sequenced after three rounds of phage library screening. The sequences showed a striking consensus (**Table 13**). The HPQ motifs were predominately adjacent to the C-terminus of the "loop" region. This differed from the linear peptides that were found to bind streptavidin, in which the HPQ

motif could be found in any position within the peptides. This consensus suggests that the **xxtz1** scaffold is structured when random peptide sequences were inserted, because a similar consensus was also observed among cyclic peptides [92].

5.3.5 Characterization of streptavidin binding proteins

To test whether the constrained **xxtz1** scaffold indeed improved the peptide affinity to streptavidin, one of the peptide sequences was chosen to be characterized for its affinity to streptavidin by surface plasmon resonance (SPR) (**Table 13**). The signal response in SPR is directly related to the size of the species that is binding to the surface; as a result, the peptide sequence was genetically inserted between CFP and YFP to increase SPR response. The resulting protein was designated as **xxtz1-HPQ1** (**Table 13**). Pre-streptavidin-coated chip from Biacore was used. Injecting increasing concentrations of the FRET constructs led to corresponding increases in the sensorgram signal (**Figure 38**). The **xxtz1-HPQ1** dissociation constant (K_d) to streptavidin was calculated as 320 nM, which agrees with a previously-reported cyclic peptide bearing the HPQ motif [200].

As a comparison, a "linear" version **xxtz1-HPQ1** was also tested by SPR. This non-structured version was created by replacing the structure-dependent amino acids at the position 3, 5, 7, 9, 16, 18, 20, and 22 in the **xxtz1** peptide with non-structurally-constrained glycine. The rest of the **xxtz1** peptide sequence and the streptavidin binding sequence EMQDHPQ are the same as **xxtz1-HPQ1**. This glycine version peptide was also fused between CFP and YFP and subjected to identical SPR procedures for K_d measurement. A much lower response in SPR was noticed on this non-structured version (**Figure 38**). The calculated K_d is 130 μM, 3 orders of magnitude lower than the structured counterpart. This result agrees with what was previously found for the linear peptides that bind to streptavidin [201]. These results showed that **xxtz1** "stem" scaffold served to preorganize the HPQ-containing loop for the high affinity streptavidin recognition.

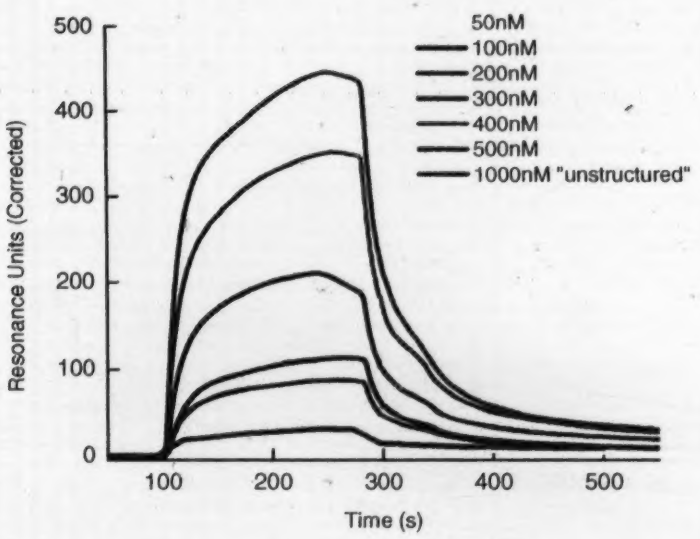


Figure 38 Sensorgrams of the binding of **xxtz1-HPQ1** fusion proteins to a streptavidin coating surface. The **xxtz1-HPQ1** was injected at concentration ranging from 50 nM to 500 nM, and the binding affinity was determined by analysis of the on- and off- rates. The "unstructured" counterpart was also tested. The "unstructured" version was obtained by replacing the structure-dependent amino acids with glycines. The sensor chip was regenerated with 50 mM NaOH after each binding experiment.

5.4 CONCLUSION

In an effort to develop a structured peptide scaffold that lacks a disulfide bond and is thus suitable for molecular recognition applications in the reducing environment of the cytosol, we have investigated engineered versions of the trpzip. Our early results suggested that even the most highly-folded member of the trpzip class (the 16-mer **HP5W4**), were substantially destabilized by the presence of mutations in the turn region and therefore was not an ideal peptide scaffold. To address this issue, we used a fluorescence-based live-cell screening system to identify extended trpzip-type variants with additional stabilizing interactions. One of these extended trpzip-type variants is the 24-mer **xxtz1**-peptide with the sequence KAWTHDWTWNPATGKWTWLWRKNK.

Interestingly, we found that in the context of a fusion with two fluorescent proteins, **xxtz1**-peptide seemed to remain highly structured when a 7-residue unstructured loop was inserted into the turn region. The resulting peptide was envisioned to adapt a stem-loop-type structure.

To test the ability of **xxtz1**-peptide to serve as a scaffold for molecular recognition, randomized peptide libraries were grafted into the loop region by standard molecular biology techniques and displayed on the surface of filamentous phage M13 as a low copy number pIII fusion. Panning of the resulting libraries against immobilized streptavidin resulted in a dramatic enrichment of phage displaying peptide sequences with submicromolar affinities for streptavidin. Mutating key residues in the "stem" portion of the peptide structure resulted in a 3 order-of-magnitude drop in K_d, illustrating the critical importance of the stem portion in preorganizing the loop region for molecular recognition.

Overall, these results provide a tantalizing hint that it may indeed be possible to engineer stable peptide-based stem-loop structures that are highly soluble and highly folded in solution. Our **xxtz1(loop)**-peptide is limited by the fact that it seems to be poorly folded in solution and highly folded only in the context of the CFP-YFP fusion protein. Evidence for this structure include: high FRET in the context of the CFP-YFP fusion protein; the bias of the HPQ sequence to the C-terminal end of the loop (as observed in covalently cyclized peptides); and the critical dependence of high-affinity binding on the Trp residues of the "stem" portion of the stem-loop structure. On the other hand, the diminished strength of the 227 nm exciton in the CD spectrum indicates that the free peptide is only about 30% folded. A future direction for this project might be to directly screen libraries of stem-loop-type structures at elevated temperatures to identify more promising non-cyclized generic peptide scaffolds for molecular recognition applications.

CHAPTER 6:

CONCLUSIONS AND FUTURE DIRECTIONS

It is well known that protein (Greek: proteios, "primary important" [202]) plays a central role in essentially all biological processes. The apparent versatility of proteins suggests that they could be extremely powerful molecular tools if only researchers could change an existing protein or create entirely new ones with desired functions. One obvious application of proteins with unnatural functions would be catalysts for reactions of industrial or biomedical interest. Another would be binding proteins with molecular recognition properties of interest to the biomedical community. With the advent of molecular biology techniques, protein engineering has become a productive and popular research discipline. To date, numerous proteins with desired properties for different application purposes have been engineered. One milestone in this field is the engineering of non-immunoglobulin proteins for specific molecular recognition. The non-immunoglobulin domains, best suited for conversion into molecular recognition domains, have been referred to as "generic protein scaffolds" [74, 140]. Desirable features of a generic protein scaffold for intracellular applications include a relatively small size and the absence of disulfide bonds. In an effort to add a new mini-motif to the growing repertoire of validated non-immunoglobulin binding proteins, we have undertaken the development of stable hairpin peptides that can withstand randomized amino-acids substitutions and insertions. As I have demonstrated in this body of work, these stable peptides have proven themselves to be promising scaffolds for molecular recognition.

6.1 FRET APPROACH IN PROTEIN ENGINEERING

We have developed and validated a FRET-based library-screening method for identifying hairpin peptide variants with stable structures. This FRET strategy can be used both *in vitro* and *in vivo*. The two identified stable variants, **xtz1** and **xxtz1**, were subsequently used as scaffolds for streptavidin recognition.

It is reasonable to suggest that further trpzip optimization could be carried out by this FRET-based approach. It will be worthwhile to apply this strategy to trpzip peptides with randomization of residues that were not investigated in this study (*i.e.*, residues in the

turn region) to identify highly structural variants. Furthermore, it would be interesting to continue applying our "extend" strategy to the **xxtz1**-peptide, in the hope of indentifying even longer hairpin variants. Presumably, longer variants would better tolerate the introduction of destabilizing mutations as would be present in a peptide library. One limitation in this FRET approach is that screening is based on the distance (or locked orientation) between the peptide termini rather than the overall thermal stability. All libraries screening in this work were done at room temperature. Alternatively, library screening at high temperature could result in the identification of more thermally stable peptide structures. Variants that are better able to retain their structure at elevated temperature would provide a higher FRET signal. Based on the high stability of FPs, it may be possible to perform this screen at temperatures higher than 60 °C. Of course, there would be technical challenges associated with warming plates of *E. coli* colonies to these temperatures.

Although my efforts were focused on tryptophan zipper β-hairpin peptides, I do expect that this high-throughput screening platform technology could be well adapted for screening other small genetically-encoded peptide motifs for folding efficiency or stability. For intracellular applications, the scaffold resistance to proteolysis is also very important. This FRET approach could also be employed to screen variants with better protease resistance. A dual expression vector, in which the FRET constructs and protease(s) are under the control of different promoters, could be used in this case. Once the protease(s) is activated, a decrease in FRET signal would be noticed (this strategy was discussed in the **Chapter 1, Figure 7B**). The genetically-inserted peptide/protein variants ability to resist protease(s) activity could be revealed from the rate of FRET decrease.

6.2 TRPZIP VARIANTS SCAFFOLDS

The **xtz1** and **xxtz1** trpzip variants show great potential as generic protein scaffolds. The strategy of substituting randomized amino acids into the non-Trp-face residues in the

xtz1 is essentially similar to the strategy reported for the Zinc finger motif [109] and Trp cage motif [105] (**Chapter 1**), in which the stable secondary structure is critical to the high affinity binding. The second strategy, randomized loop insertion, provides more binding site conformation flexibility. Generally, this strategy is more effective at generating "universal" binders. One good example in this class is the cyclic peptide scaffold. Hundreds of these constrained peptides have been reported that can bind a variety of targets, such as viruses [203], enzymes [204], even metal crystals [205]. A commercial cyclic peptide phage display kit is available (New England Biolabs). However, owing to the disulfide bond, cyclic peptides cannot be used for intracellular applications. Fortunately, the **xxtz1** scaffold overcomes this limitation. An obvious future direction of this research is to screen for intracellular component recognition using the trpzip variant-based scaffolds. Once identified, they will be invaluable tools for live cell imaging as well as potential disease diagnostic reagents. In the case of only low affinity binders, a tandem-copy strategy could be employed to increase the affinity (discussed in **Chapter 1** and **2**).

Another very interesting potential application of the trpzip variants is to graft them into the FPs. FPs have a unique β-barrel structure which is made up of eleven β-strands. As mentioned in **Chapter 1**, avGFP has been was used as a scaffold for randomized peptide(s) presentation. But this strategy suffered from avGFP's β-strands stability in tolerating loop insertions. Using protein engineering techniques, variants of trpzip bearing preorganized molecular recognition loops could potentially replace one or several of the β-strands of the FPs, while not severely diminishing their fluorescence. Thus, the resulting protein(s) would not only be a fluorescent probes but also a binding molecule suitable for intracellular imaging.

BIBLIOGRAPHY

1. Cheng, Z., and Campbell, R.E., Assessing the structural stability of designed beta-hairpin peptides in the cytoplasm of live cells. *ChemBioChem*, 2006. **7**(8): p. 1147-50.
2. Cheng, Z., Miskolzie, M., and Campbell, R.E., In vivo screening identifies a highly folded beta-hairpin peptide with a structured extension. *Chembiochem*, 2007. **8**(8): p. 880-3.
3. Cheng, Z. and Campbell, R.E., Fluorescence-based characterization of genetically encoded peptides that fold in live cells: progress towards a generic hairpin scaffold. *SPIE* (Eds.: S. Achilefu, D. J. Bornhop, R. Raghavachari, A. P. Savitsky, R. M. Wachter), 2007. **6449**: p. 66490S.
4. Tsien, R.Y., The green fluorescent protein. *Annu. Rev. Biochem.*, 1998. **67**: p. 509-44.
5. Shimomura, O., Johnson, F.H., and Saiga, Y., Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.*, 1962. **59**(3): p. 223-239.
6. Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J., Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 1992. **111**(2): p. 229-33.
7. Morise, H., Shimomura, O., Johnson, F.H., and Winant, J., Intermolecular energy transfer in the bioluminescent system of Aequorea. *Biochemistry*, 1974. **13**(12): p. 2656-2662.
8. Ormo, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J., Crystal structure of the Aequorea victoria green fluorescent protein. *Science*, 1996. **273**(5280): p. 1392-5.
9. Yang, F., Moss, L.G., and Phillips, G.N., Jr., The molecular structure of green fluorescent protein. *Nat. Biotechnol.*, 1996. **14**(10): p. 1246-51.

10. Cubitt, A.B., Heim, R., Adams, S.R., Boyd, A.E., Gross, L.A., and Tsien, R.Y., Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci*,, 1995. **20**(11): p. 448-55.

11. Heim, R., Cubitt, A.B., and Tsien, R.Y., Improved green fluorescence. *Nature*, 1995. **373**(6516): p. 663-4.

12. Heim, R., Prasher, D.C., and Tsien, R.Y., Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.*, 1994. **91**(26): p. 12501-4.

13. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C., Green fluorescent protein as a marker for gene expression. *Science*, 1994. **263**(5148): p. 802-5.

14. Ehrig, T., O'Kane, D.J., and Prendergast, F.G., Green-fluorescent protein mutants with altered fluorescence excitation spectra. *FEBS Lett.*, 1995. **367**(2): p. 163-6.

15. Heim, R. and Tsien, R.Y., Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.*, 1996. **6**(2): p. 178-82.

16. Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A., A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat. Biotechnol.*, 2002. **20**(1): p. 87-90.

17. Zacharias, D.A., Violin, J.D., Newton, A.C., and Tsien, R.Y., Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*, 2002. **296**(5569): p. 913-6.

18. Delagrave, S., Hawtin, R.E., Silva, C.M., Yang, M.M., and Youvan, D.C., Red-shifted excitation mutants of the green fluorescent protein. *Biotechnology (N Y)*, 1995. **13**(2): p. 151-4.

19. Cormack, B.P., Valdivia, R.H., and Falkow, S., FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 1996. **173**(1): p. 33-38.

20. Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P., Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat. Biotechnol.*, 1996. **14**(3): p. 315-9.

21. Volkov, A.A., Shao, Z., and Arnold, F.H., Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair. *Nucleic Acids Res.*, 1999. **27**(18): p. e18.

22. Llopis, J., McCaffery, J.M., Miyawaki, A., Farquhar, M.G., and Tsien, R.Y., Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 1998. **95**(12): p. 6803-8.

23. Miyawaki, A., Llopis, J., Heim, R., McCaffery, J.M., Adams, J.A., Ikura, M., and Tsien, R.Y., Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature*, 1997. **388**(6645): p. 882-7.

24. Romoser, V.A., Hinkle, P.M., and Persechini, A., Detection in living cells of Ca2+-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. A new class of fluorescent indicators. *J. Biol. Chem.*, 1997. **272**(20): p. 13270-4.

25. Knight, C.G., Fluorimetric assays of proteolytic enzymes. *Methods Enzymol.*, 1995. **248**: p. 18-34.

26. Cubitt, A.B., Woollenweber, L.A., and Heim, R., Understanding structure-function relationships in the Aequorea victoria green fluorescent protein. *Methods Cell Biol.*, 1999. **58**: p. 19-30.

27. Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R., and Piston, D.W., Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.*, 1997. **73**(5): p. 2782-90.

28. Ai, H.W., Shaner, N.C., Cheng, Z., Tsien, R.Y., and Campbell, R.E., Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry*, 2007. **46**(20): p. 5904-10.

29. Wachter, R.M., Elsliger, M.A., Kallio, K., Hanson, G.T., and Remington, S.J., Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure*, 1998. **6**(10): p. 1267-77.

30. Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A., Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol.*, 1999. **17**(10): p. 969-73.

31. Wall, M.A., Socolich, M., and Ranganathan, R., The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Biol.*, 2000. **7**(12): p. 1133-8.

32. Yarbrough, D., Wachter, R.M., Kallio, K., Matz, M.V., and Remington, S.J., Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-A resolution. *Proc. Natl. Acad. Sci. U.S.A.*, 2001. **98**(2): p. 462-7.

33. Gross, L.A., Baird, G.S., Hoffman, R.C., Baldridge, K.K., and Tsien, R.Y., The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U.S.A.*, 2000. **97**(22): p. 11990-5.

34. Forster, T., Intermolecular energy migration and fluorescence. *Ann. Phys. (Leipzig)*, 1948. **2**: p. 55-75.

35. dos Remedios, C.G. and Moens, P.D., Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. *J. Struct. Biol.*, 1995. **115**(2): p. 175-85.

36. Van Der Meer, B.W., Coker, G.III, and Chen, S.Y., Resonance Energy Transfer: Theory and Data. 1994, New York: VCH.

37. Stryer, L. and Haugland, R.P., Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U.S.A.*, 1967. **58**(2): p. 719-26.

38. Fradkov, A.F., Verkhusha, V.V., Staroverov, D.B., Bulina, M.E., Yanushevich, Y.G., Martynov, V.I., Lukyanov, S., and Lukyanov, K.A., Far-red fluorescent tag for protein labelling. *Biochem. J.*, 2002. **368**(Pt 1): p. 17-21.

39. Tramier, M., Zahid, M., Mevel, J.C., Masse, M.J., and Coppey-Moisan, M., Sensitivity of CFP/YFP and GFP/mCherry pairs to donor photobleaching on FRET determination by fluorescence lifetime imaging microscopy in living cells. *Microsc. Res. Tech.*, 2006. **69**(11): p. 933-9.

40. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E., and Tsien, R.Y., Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.*, 2004. **22**(12): p. 1567-72.

41. Laffray, S., Tan, K., Dulluc, J., Bouali-Benazzouz, R., Calver, A.R., Nagy, F., and Landry, M., Dissociation and trafficking of rat GABAB receptor heterodimer upon chronic capsaicin stimulation. *Eur. J. Neurosci.*, 2007. **25**(5): p. 1402-16.

42. Yang, X., Xu, P., and Xu, T., A new pair for inter- and intra-molecular FRET measurement. *Biochem. Biophys. Res. Commun.*, 2005. **330**(3): p. 914-20.

43. Peter, M., Ameer-Beg, S.M., Hughes, M.K., Keppler, M.D., Prag, S., Marsh, M., Vojnovic, B., and Ng, T., Multiphoton-FLIM quantification of the EGFP-mRFP1 FRET pair for localization of membrane receptor-kinase interactions. *Biophys. J.*, 2005. **88**(2): p. 1224-37.

44. Kawai, H., Suzuki, T., Kobayashi, T., Ishii-Watabe, A., Sakurai, H., Ohata, H., Honda, K., Momose, K., Hayakawa, T., and Kawanishi, T., Caspase cascade proceeds rapidly after cytochrome c release from mitochondria in tumor necrosis factor-alpha-induced cell death. *J. Pharmacol. Sci.*, 2007. **103**(2): p. 159-67.

45. Elphick, L.M., Meinander, A., Mikhailov, A., Richard, M., Toms, N.J., Eriksson, J.E., and Kass, G.E., Live cell detection of caspase-3 activation by a Discosoma-red-fluorescent-protein-based fluorescence resonance energy transfer construct. *Anal. Biochem.*, 2006. **349**(1): p. 148-55.

46. Hsu, Y.Y., Liu, Y.N., Wang, W., Kao, F.J., and Kung, S.H., In vivo dynamics of enterovirus protease revealed by fluorescence resonance emission transfer (FRET) based on a novel FRET pair. *Biochem. Biophys. Res. Commun.*, 2007. **353**(4): p. 939-45.

47. Mizuno, H., Sawano, A., Eli, P., Hama, H., and Miyawaki, A., Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer. *Biochemistry*, 2001. **40**(8): p. 2502-10.

48. Erickson, M.G., Moon, D.L., and Yue, D.T., DsRed as a potential FRET partner with CFP and GFP. *Biophys. J.*, 2003. **85**(1): p. 599-611.

49. Souslova, E.A. and Chudakov, D.M., Photoswitchable cyan fluorescent protein as a FRET donor. *Microsc. Res. Tech.*, 2006. **69**(3): p. 207-9.

50. Ganesan, S., Ameer-Beg, S.M., Ng, T.T., Vojnovic, B., and Wouters, F.S., A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACh) for Forster resonance energy transfer with GFP. *Proc. Natl. Acad. Sci. U.S.A.*, 2006. **103**(11): p. 4089-94.

51. Schleifenbaum, A., Stier, G., Gasch, A., Sattler, M., and Schultz, C., Genetically encoded FRET probe for PKC activity based on pleckstrin. *J. Am. Chem. Soc.*, 2004. **126**(38): p. 11786-7.

52. Rizzo, M.A., Springer, G., Segawa, K., Zipfel, W.R., and Piston, D.W., Optimization of pairings and detection conditions for measurement of FRET between cyan and yellow fluorescent proteins. *Microsc. Microanal.*, 2006. **12**(3): p. 238-54.

53. Kremers, G.J., Goedhart, J., van Munster, E.B., and Gadella, T.W., Jr., Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Forster radius. *Biochemistry*, 2006. **45**(21): p. 6570-80.

54. Nguyen, A.W. and Daugherty, P.S., Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.*, 2005. **23**(3): p. 355-60.

55. Patterson, G.H., Piston, D.W., and Barisas, B.G., Forster Distances between Green Fluorescent Protein Pairs. *Anal. Biochem.*, 2000. **284**(2): p. 438-440.

56. Kogure, T., Karasawa, S., Araki, T., Saito, K., Kinjo, M., and Miyawaki, A., A fluorescent variant of a protein from the stony coral Montipora facilitates dual-color single-laser fluorescence cross-correlation spectroscopy. *Nat. Biotechnol.*, 2006. **24**(5): p. 577-81.

57. Ward, W.W., Biochemical and physical properties of GFP, in Green Fluorescent Protein: Properties, Applications, and Protocols, M.K. Chalfie, S., Editor. 1998, Wiley: New York. p. 45-75.

58. Mena, M.A., Treynor, T.P., Mayo, S.L., and Daugherty, P.S., Blue fluorescent proteins with enhanced brightness and photostability from a structurally targeted library. *Nat. Biotechnol.*, 2006. **24**(12): p. 1569-71.

59. Ai, H.W., Henderson, J.N., Remington, S.J., and Campbell, R.E., Directed evolution of a monomeric, bright and photostable version of Clavularia cyan fluorescent protein: structural characterization and applications in fluorescence imaging. *Biochem. J.*, 2006. **400**(3): p. 531-40.

60. Shaner, N.C., Patterson, G.H., and Davidson, M.W., Advances in fluorescent protein technology. *J. Cell Sci.*, 2007. **120**(Pt 24): p. 4247-60.

61. Ai, H.-w., Hazelwood, K.L., Davidson, M.W., and Campbell, R.E., Fluorescent protein FRET pairs for ratiometric imaging of dual biosensors. *Nat. Methods*, 2008. **5**:p. 401-3.

62. Janetopoulos, C., Jin, T., and Devreotes, P., Receptor-Mediated Activation of Heterotrimeric G-Proteins in Living Cells. *Science*, 2001. **291**(5512): p. 2408-2411.

63. Tertoolen, L., Blanchetot, C., Jiang, G., Overvoorde, J., Gadella, T., Hunter, T., and Hertog, J.d., Dimerization of Receptor Protein-Tyrosine Phosphatase alpha in living cells. *BMC Cell Biology*, 2001. **2**(1): p. 8.

64. Llopis, J., Westin, S., Ricote, M., Wang, Z., Cho, C.Y., Kurokawa, R., Mullen, T.M., Rose, D.W., Rosenfeld, M.G., Tsien, R.Y., and Glass, C.K., Ligand-dependent interactions of coactivators steroid receptor coactivator-1 and peroxisome proliferator-activated receptor binding protein with nuclear hormone receptors can be imaged in live cells and are required for transcription. *Proc. Natl. Acad. Sci. U.S.A.*, 2000. **97**(8): p. 4363-8.

65. Vanderklish, P.W., Krushel, L.A., Holst, B.H., Gally, J.A., Crossin, K.L., and Edelman, G.M., Marking synaptic activity in dendritic spines with a calpain substrate exhibiting fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U.S.A.*, 2000. **97**(5): p. 2253-8.

66. Winter, G. and Milstein, C., Man-made antibodies. *Nature*, 1991. **349**(6307): p. 293-9.

67. Michaud, G.A., Salcius, M., Zhou, F., Bangham, R., Bonin, J., Guo, H., Snyder, M., Predki, P.F., and Schweitzer, B.I., Analyzing antibody specificity with whole proteome microarrays. *Nat. Biotechnol.*, 2003. **21**(12): p. 1509-12.

68. Carter, P.J., Potent antibody therapeutics by design. *Nat. Rev. Immunol.*, 2006. **6**(5): p. 343-357.

69. Smith, G.P. and Petrenko, V.A., Phage Display. *Chem. Rev.*, 1997. **97**(2): p. 391-410.

70. Roberts, R.W. and Szostak, J.W., RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 1997. **94**(23): p. 12297-302.

71. Hanes, J. and Pluckthun, A., In vitro selection and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. U.S.A.*, 1997. **94**(10): p. 4937-42.

72. Georgiou, G., Stephens, D.L., Stathopoulos, C., Poetschke, H.L., Mendenhall, J., and Earhart, C.F., Display of beta-lactamase on the Escherichia coli surface: outer membrane phenotypes conferred by Lpp'-OmpA'-beta-lactamase fusions. *Protein. Eng.*, 1996. **9**(2): p. 239-47.

73. Samuelson, P., Gunneriusson, E., Nygren, P.A., and Stahl, S., Display of proteins on bacteria. *J. Biotechnol.*, 2002. **96**(2): p. 129-54.

74. Skerra, A., Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.*, 2000. **13**(4): p. 167-87.

75. Holliger, P., and Hudson, P.J., Engineered antibody fragments and the rise of single domain. *Nat. Biotechnol.*, 2005. 23(9): p.a1126-36.

76. Binz, H.K., Amstutz, P., and Pluckthun, A., Engineering novel binding proteins from nonimmunoglobulin domains. *Nat. Biotechnol.*, 2005. **23**(10): p. 1257-68.

77. de Graaf, M., van der Meulen-Muileman, I.H., Pinedo, H.M., and Haisma, H.J., Expression of scFvs and scFv fusion proteins in eukaryotic cells. *Methods Mol. Biol.*, 2002. **178**: p. 379-87.

78. De Genst, E., Handelberg, F., Van Meirhaeghe, A., Vynck, S., Loris, R., Wyns, L., and Muyldermans, S., Chemical basis for the affinity maturation of a camel single domain antibody. *J. Biol. Chem.*, 2004. **279**(51): p. 53593-601.

79. Streltsov, V.A., Varghese, J.N., Carmichael, J.A., Irving, R.A., Hudson, P.J., and Nuttall, S.D., Structural evidence for evolution of shark Ig new antigen receptor

variable domain antibodies from a cell-surface receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 2004. **101**(34): p. 12444-9.

80. Dooley, H. and Flajnik, M.F., Shark immunity bites back: affinity maturation and memory response in the nurse shark, Ginglymostoma cirratum. *Eur. J. Immunol.*, 2005. **35**(3): p. 936-45.

81. Muyldermans, S., Single domain camel antibodies: current status. *Rev. Mol. Biotechnol.*, 2001. **74**(4): p. 277-302.

82. Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T.A., and Uhlen, M., A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.*, 1987. **1**(2): p. 107-13.

83. Nord, K., Nilsson, J., Nilsson, B., Uhlen, M., and Nygren, P.A., A combinatorial library of an alpha-helical bacterial receptor domain. *Protein Eng.*, 1995. **8**(6): p. 601-8.

84. Sandstrom, K., Xu, Z., Forsberg, G., and Nygren, P.A., Inhibition of the CD28-CD80 co-stimulation signal by a CD28-binding affibody ligand developed by combinatorial protein engineering. *Protein Eng.*, 2003. **16**(9): p. 691-7.

85. Gunneriusson, E., Nord, K., Uhlen, M., and Nygren, P., Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling. *Protein Eng.*, 1999. **12**(10): p. 873-8.

86. Nord, K., Gunneriusson, E., Uhlen, M., and Nygren, P.A., Ligands selected from combinatorial libraries of protein A for use in affinity capture of apolipoprotein A-1M and taq DNA polymerase. *J. Biotechnol.*, 2000. **80**(1): p. 45-54.

87. Orlova, A., Magnusson, M., Eriksson, T.L., Nilsson, M., Larsson, B., Hoiden-Guthenberg, I., Widstrom, C., Carlsson, J., Tolmachev, V., Stahl, S., and Nilsson, F.Y., Tumor imaging using a picomolar affinity HER2 binding affibody molecule. *Cancer Res.*, 2006. **66**(8): p. 4339-48.

88. Orlova, A., Tolmachev, V., Pehrson, R., Lindborg, M., Tran, T., Sandstrom, M., Nilsson, F.Y., Wennborg, A., Abrahmsen, L., and Feldwisch, J., Synthetic affibody molecules: a novel class of affinity ligands for molecular imaging of HER2-expressing malignant tumors. *Cancer Res.*, 2007. **67**(5): p. 2178-86.

89. Tolmachev, V., Orlova, A., Pehrson, R., Galli, J., Baastrup, B., Andersson, K., Sandstrom, M., Rosik, D., Carlsson, J., Lundqvist, H., Wennborg, A., and Nilsson, F.Y., Radionuclide therapy of HER2-positive microxenografts using a 177Lu-labeled HER2-specific Affibody molecule. *Cancer Res.*, 2007. **67**(6): p. 2773-82.

90. Steffen, A.C., Gostring, L., Tolmachev, V., Palm, S., Stenerlow, B., and Carlsson, J., Differences in radiosensitivity between three HER2 overexpressing cell lines. *Eur. J. Nucl. Med. Mol. Imaging*, 2008.

91. Chak, K.F., Safo, M.K., Ku, W.Y., Hsieh, S.Y., and Yuan, H.S., The crystal structure of the immunity protein of colicin E7 suggests a possible colicin-interacting surface. *Proc. Natl. Acad. Sci. U.S.A.*, 1996. **93**(13): p. 6552-6.

92. Ku, J., and Schultz, P.G., Alternate protein frameworks for molecular recognition. *Proc. Natl. Acad. Sci. U.S.A.*, 1995. **92**(14): p. 6437-6442.

93. Landschulz, W.H., Johnson, P.F., and McKnight, S.L., The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, 1988. **240**(4860): p. 1759-64.

94. Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Pluckthun A., Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.*, 2003. **332**(2): p. 489-503.

95. Pancer, Z., Amemiya, C.T., Ehrhardt, G.R.A., Ceitlin, J., Larry Gartland, G., and Cooper, M.D., Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature*, 2004. **430**(6996): p. 174-180.

96. Gelly, J.C., Gracy, J., Kaas, Q., Le-Nguyen, D., Heitz, A., and Chiche, L., The KNOTTIN website and database: a new information system dedicated to the knottin scaffold. *Nucleic Acids Res.*, 2004. **32**(Database issue): p. D156-9.

97. Schlehuber, S. and Skerra, A., Lipocalins in drug discovery: from natural ligand-binding proteins to "anticalins". *Drug Discov. Today*, 2005. **10**(1): p. 23-33.

98. Koide, A., Bailey, C.W., Huang, X., and Koide, S., The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.*, 1998. **284**(4): p. 1141-1151.

99. Min, J., Zhang, X., Cheng, X., Grewal, S.I., and Xu, R.M., Structure of the SET domain histone lysine methyltransferase Clr4. *Nat. Struct. Biol.*, 2002. **9**(11): p. 828-32.

100. Pessi, A., Bianchi, E., Crameri, A., Venturini, S., Tramontano, A., and Sollazzo, M., A designed metal-binding protein with a novel fold. *Nature*, 1993. **362**(6418): p. 367-369.

101. Abedi, M.R., Caponigro, G., and Kamb, A., Green fluorescent protein as a scaffold for intracellular presentation of peptides. *Nucleic Acids Res.*, 1998. **26**(2): p. 623-30.

102. Matthias Paschke, C.T.W.H., Engineering a circularly permuted GFP scaffold for peptide presentation. *J. Mol. Recognit.*, 2007. **20**(5): p. 367-378.

103. Zeytun, A., Jeromin, A., Scalettar, B.A., Waldo, G.S., and Bradbury, A.R.M., Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies. *Nat. Biotechnol.*, 2003. **21**(12): p. 1473-1479.

104. Zeytun, A., Jeromin, A., Scalettar, B.A., Waldo, G.S., and Bradbury, A.R.M., Fluorobodies combine GFP fluorescence with the binding characteristics of antibodies (vol 21, pg 1473, 2003). *Nat. Biotechnol.*, 2004. **22**(5): p. 601-601.

105. Herman, R.E., Badders, D., Fuller, M., Makienko, E.G., Houston, M.E., Jr., Quay, S.C., and Johnson, P.H., The Trp cage motif as a scaffold for the display of a randomized peptide library on bacteriophage T7. *J. Biol. Chem.*, 2007. **282**(13): p. 9813-24.

106. Bianchi, E., Folgori, A., Wallace, A., Nicotra, M., Acali, S., Phalipon, A., Barbato, G., Bazzo, R., Cortese, R., Felici, F., and et al., A conformationally homogeneous combinatorial peptide library. *J. Mol. Biol.*, 1995. **247**(2): p. 154-60.

107. Ali, S.A., Joao, H.C., Hammerschmid, F., Eder, J., and Steinkasserer, A., Transferrin trojan horses as a rational approach for the biological delivery of therapeutic peptide domains. *J. Biol. Chem.*, 1999. **274**(34): p. 24066-73.

108. Legendre, D., Vucic, B., Hougardy, V., Girboux, A.L., Henrioul, C., Van Haute, J., Soumillion, P., and Fastrez, J., TEM-1 beta-lactamase as a scaffold for protein recognition and assay. *Protein Sci.*, 2002. **11**(6): p. 1506-18.

109. Klug, A., Zinc finger peptides for the regulation of gene expression. *J. Mol. Biol.*, 1999. **293**(2): p. 215-8.

110. Brown, R.S., Sander, C., and Argos, P., The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.*, 1985. **186**(2): p. 271-4.

111. Miller, J., McLachlan, A.D., and Klug, A., Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *Embo. J.*, 1985. **4**(6): p. 1609-14.

112. Choo, Y. and Klug, A., Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc. Natl. Acad. Sci. U.S.A.*, 1994. **91**(23): p. 11168-72.

113. Jamieson, A.C., Kim, S.H., and Wells, J.A., In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry*, 1994. **33**(19): p. 5689-95.

114. Rebar, E.J. and Pabo, C.O., Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, 1994. **263**(5147): p. 671-3.

115. Wu, H., Yang, W.P., and Barbas, C.F., 3rd, Building zinc fingers by selection: toward a therapeutic application. *Proc. Natl. Acad. Sci. U.S.A.*, 1995. **92**(2): p. 344-8.

116. Beerli, R.R., Segal, D.J., Dreier, B., and Barbas, C.F., 3rd, Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl. Acad. Sci. U.S.A.*, 1998. **95**(25): p. 14628-33.

117. Neidigh, J.W., Fesinmeyer, R.M., Prickett, K.S., and Andersen, N.H., Exendin-4 and glucagon-like-peptide-1: NMR structural comparisons in the solution and micelle-associated states. *Biochemistry*, 2001. **40**(44): p. 13188-200.

118. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., and Yanagawa, H., DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res.*, 2003. **31**(19): p. e118.

119. Speight, R.E., Hart, D.J., Sutherland, J.D., and Blackburn, J.M., A new plasmid display technology for the in vitro selection of functional phenotype-genotype linked proteins. *Chem. Biol.*, 2001. **8**(10): p. 951-65.

120. Terry, T.D., Malik, P., and Perham, R.N., Accessibility of peptides displayed on filamentous bacteriophage virions: susceptibility to proteinases. *Biol. Chem.*, 1997. **378**(6): p. 523-30.

121. Makowski, L., Terminating a macromolecular helix. Structural model for the minor proteins of bacteriophage M13. *J. Mol. Biol.*, 1992. **228**(3): p. 885-92.

122. Stengele, I., Bross, P., Garces, X., Giray, J., and Rasched, I., Dissection of functional domains in phage fd adsorption protein. Discrimination between attachment and penetration sites. *J. Mol. Biol.*, 1990. **212**(1): p. 143-9.

123. Webster, R.E., The tol gene products and the import of macromolecules into Escherichia coli. *Mol. Microbiol.*, 1991. **5**(5): p. 1005-11.

124. Barbas, C.F., 3rd, Burton, D.R., Scott, J.K., and Silverman, G.J., Phage Display: a Laboratory Manual. 2001: Cold Spring Harbor Laboratory Press.

125. Gray, C.W., Three-dimensional structure of complexes of single-stranded DNA-binding proteins with DNA. IKe and fd gene 5 proteins form left-handed helices with single-stranded DNA. *J. Mol. Biol.*, 1989. **208**(1): p. 57-64.

126. Kazmierczak, B.I., Mielke, D.L., Russel, M., and Model, P., pIV, a filamentous phage protein that mediates phage export across the bacterial cell envelope, forms a multimer. *J. Mol. Biol.*, 1994. **238**(2): p. 187-98.

127. Rapoza, M.P. and Webster, R.E., The filamentous bacteriophage assembly proteins require the bacterial SecA protein for correct localization to the membrane. *J. Bacteriol.*, 1993. **175**(6): p. 1856-9.

128. Lopez, J. and Webster, R.E., Morphogenesis of filamentous bacteriophage f1: orientation of extrusion and production of polyphage. *Virology*, 1983. **127**(1): p. 177-93.

129. Iannolo, G., Minenkova, O., Petruzzelli, R., and Cesareni, G., Modifying filamentous phage capsid: limits in the size of the major capsid protein. *J. Mol. Biol.*, 1995. **248**(4): p. 835-44.

130. Malik, P., Terry, T.D., Gowda, L.R., Langara, A., Petukhov, S.A., Symmons, M.F., Welsh, L.C., Marvin, D.A., and Perham, R.N., Role of capsid structure and membrane protein processing in determining the size and copy number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.*, 1996. **260**(1): p. 9-21.

131. Bass, S., Greene, R., and Wells, J.A., Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins*, 1990. **8**(4): p. 309-14.

132. Smith, G.P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science*, 1985. **228**(4705): p. 1315-7.

133. Fernandez-Gacio, A., Uguen, M., and Fastrez, J., Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol.*, 2003. **21**(9): p. 408-14.

134. Marasco, W.A., Intracellular antibodies (intrabodies) as research reagents and therapeutic molecules for gene therapy. *Immunotechnology*, 1995. **1**(1): p. 1-19.

135. Kontermann, R.E., Intrabodies as therapeutic agents. *Methods*, 2004. **34**(2): p. 163-70.

136. Nizak, C., Martin-Lluesma, S., Moutel, S., Roux, A., Kreis, T.E., Goud, B., and Perez, F., Recombinant antibodies against subcellular fractions used to track endogenous Golgi protein dynamics in vivo. *Traffic*, 2003. **4**(11): p. 739-53.

137. Nizak, C., Monier, S., del Nery, E., Moutel, S., Goud, B., and Perez, F., Recombinant antibodies to the small GTPase Rab6 as conformation sensors. *Science*, 2003. **300**(5621): p. 984-7.

138. Bird, R.E., Hardman, K.D., Jacobson, J.W., Johnson, S., Kaufman, B.M., Lee, S.M., Lee, T., Pope, S.H., Riordan, G.S., and Whitlow, M., Single-chain antigen-binding proteins. *Science*, 1988. **242**(4877): p. 423-6.

139. Glockshuber, R., Schmidt, T., and Pluckthun, A., The disulfide bonds in antibody variable domains: effects on stability, folding in vitro, and functional expression in Escherichia coli. *Biochemistry*, 1992. **31**(5): p. 1270-9.

140. Nygren, P.A. and Skerra, A., *Binding proteins from alternative scaffolds.* J. Immunol. Methods, 2004. **290**(1-2): p. 3-28.

141. Gururaja, T.L., Narasimhamurthy, S., Payan, D.G., and Anderson, D.C., A novel artificial loop scaffold for the noncovalent constraint of peptides. *Chem. Biol.*, 2000. **7**(7): p. 515-27.

142. Marks, K.M., Rosinov, M., and Nolan, G.P., In vivo targeting of organic calcium sensors via genetically selected peptides. *Chem. Biol.*, 2004. **11**(3): p. 347-56.

143. Fahnestock, S.R., Alexander, P., Nagle, J., and Filpula, D., Gene for an immunoglobulin-binding protein from a group G streptococcus. *J. Bacteriol.*, 1986. **167**(3): p. 870-80.

144. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P., Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2: why small proteins tend to have high denaturation temperatures. *Biochemistry*, 1992. **31**(14): p. 3597-603.

145. Blanco, F.J., Rivas, G., and Serrano, L., A short linear peptide that folds into a native stable [beta]-hairpin in aqueous solution. *Nat. Struct. Mol. Biol.*, 1994. **1**(9): p. 584-590.

146. Cochran, A.G., Skelton, N.J., and Starovasnik, M.A., Tryptophan zippers: stable, monomeric beta -hairpins. *Proc. Natl. Acad. Sci. U.S.A.*, 2001. **98**(10): p. 5578-83.

147. Andersen, N.H., Olsen, K.A., Fesinmeyer, R.M., Tan, X., Hudson, F.M., Eidenschink, L.A., and Farazi, S.R., Minimization and optimization of designed beta-hairpin folds. *J. Am. Chem. Soc.*, 2006. **128**(18): p. 6101-6110.

148. Butterfield, S.M. and Waters, M.L., A designed beta-hairpin peptide for molecular recognition of ATP in water. *J. Am. Chem. Soc.*, 2003. **125**(32): p. 9580-1.

149. Butterfield, S.M., Goodman, C.M., Rotello, V.M., and Waters, M.L., A peptide flavoprotein mimic: flavin recognition and redox potential modulation in water by a designed beta hairpin. *Angew. Chem. Int. Ed. Engl.*, 2004. **43**(6): p. 724-7.

150. Butterfield, S.M., Cooper, W.J., and Waters, M.L., Minimalist protein design: a beta-hairpin peptide that binds ssDNA. *J. Am. Chem. Soc.*, 2005. **127**(1): p. 24-5.

151. Silverman, J., Liu, Q., Bakker, A., To, W., Duguay, A., Alba, B.M., Smith, R., Rivas, A., Li, P., Le, H., Whitehorn, E., Moore, K.W., Swimmer, C., Perlroth, V., Vogt, M., Kolkman, J., and Stemmer, W.P., Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nat. Biotechnol.*, 2005. **23**(12): p. 1556-61.

152. Skelton, N.J., Blandl, T., Russell, S.J., Starovasnik, M.A., and Cochran, A.G., beta-hairpin polypeptides by design and selection. *Spectrosc-Int. J.*, 2003. **17**(2-3): p. 213-230.

153. Russell, S.J., Blandl, T., Skelton, N.J., and Cochran, A.G., Stability of cyclic beta-hairpins: asymmetric contributions from side chains of a hydrogen-bonded cross-strand residue pair. *J. Am. Chem. Soc.*, 2003. **125**(2): p. 388-95.

154. Ciani, B., Jourdan, M., and Searle, M.S., Stabilization of beta-hairpin peptides by salt bridges: role of preorganization in the energetic contribution of weak interactions. *J. Am. Chem. Soc.*, 2003. **125**(30): p. 9038-47.

155. Fesinmeyer, R.M., Hudson, F.M., and Andersen, N.H., Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.*, 2004. **126**(23): p. 7238-43.

156. Pastor, M.T., Lopez de la Paz, M., Lacroix, E., Serrano, L., and Perez-Paya, E., Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc. Natl. Acad. Sci. U.S.A.*, 2002. **99**(2): p. 614-9.

157. Pastor, M.T., Mora, P., Ferrer-Montiel, A., and Perez-Paya, E., Design of bioactive and structurally well-defined peptides from conformationally restricted libraries. *Biopolymers*, 2004. **76**(4): p. 357-65.

158. Huang, F., Hudgins, R.R., and Nau, W.M., Primary and secondary structure dependence of peptide flexibility assessed by fluorescence-based measurement of end-to-end collision rates. *J. Am. Chem. Soc.*, 2004. **126**(50): p. 16665-75.

159. Tyagi, S. and Kramer, F.R., Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.*, 1996. **14**(3): p. 303-8.

160. Rizzo, M.A., Springer, G.H., Granada, B., and Piston, D.W., An improved cyan fluorescent protein variant useful for FRET. *Nat. Biotechnol.*, 2004. **22**(4): p. 445-9.

161. Griesbeck, O., Baird, G.S., Campbell, R.E., Zacharias, D.A., and Tsien, R.Y., Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *J. Biol. Chem.*, 2001. **276**(31): p. 29188-94.

162. Lakowicz, J.R., Principles of fluorescence spectroscopy. 2nd ed. 1999, New York: Kluwer Academic/Plenum. xxiii, 698.

163. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D., Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio-Technology*, 1988. **6**(11): p. 1321-1325.

164. Edelhoch, H., Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry*, 1967. **6**(7): p. 1948-54.

165. Baird, G.S., Zacharias, D.A., and Tsien, R.Y., Circular permutation and receptor insertion within green fluorescent proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 1999. **96**(20): p. 11241-6.

166. Minor, D.L., Jr. and Kim, P.S., Measurement of the beta-sheet-forming propensities of amino acids. *Nature*, 1994. **367**(6464): p. 660-3.

167. Philipps, B., Hennecke, J., and Glockshuber, R., FRET-based in vivo screening for protein folding and increased protein stability. *J. Mol. Biol.*, 2003. **327**(1): p. 239-49.

168. Tomoaki Matsuura, A.E.D.L.Z.A.P., Combinatorial Approaches To Novel Proteins. *ChemBioChem*, 2004. **5**(2): p. 177-182.

169. Pinto, A.L., Hellinga, H.W., and Caradonna, J.P., Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proc. Natl. Acad. Sci. U.S.A.*, 1997. **94**(11): p. 5562-7.

170. Grishina, I.B. and Woody, R.W., Contributions of tryptophan side chains to the circular dichroism of globular proteins: exciton couplets and coupled oscillators. *Faraday Discuss.*, 1994(99): p. 245-62.

171. Nixon, A.E. and Firestine, S.M., Rational and "irrational" design of proteins and their use in biotechnology. *IUBMB Life*, 2000. **49**(3): p. 181-7.

172. Magliery, T.J. and Regan, L., Combinatorial approaches to protein stability and structure. *Eur. J. Biochem.*, 2004. **271**(9): p. 1595-608.

173. Watters, A.L. and Baker, D., Searching for folded proteins in vitro and in silico. *Eur. J. Biochem.*, 2004. **271**(9): p. 1615-22.

174. Ulmschneider, J.P. and Jorgensen, W.L., Polypeptide folding using Monte Carlo sampling, concerted rotation, and continuum solvation. *J. Am. Chem. Soc.*, 2004. **126**(6): p. 1849-57.

175. Streicher, W.W. and Makhatadze, G.I., Calorimetric evidence for a two-state unfolding of the beta-hairpin peptide trpzip4. *J. Am. Chem. Soc.*, 2006. **128**(1): p. 30-1.

176. Snow, C.D., Qiu, L., Du, D., Gai, F., Hagen, S.J., and Pande, V.S., Trp zipper folding kinetics by molecular dynamics and temperature-jump spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 2004. **101**(12): p. 4077-82.

177. Bax, A. and Davis, D.G., Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson.*, 1985. **65**(2): p. 355-360.

178. Levitt, M.H., Freeman, R., and Frenkiel, T., Broad-Band Heteronuclear Decoupling. *J. Magn. Reson.*, 1982. **47**(2): p. 328-330.

179. Hurd, R.E., Gradient-Enhanced Spectroscopy. *J. Magn. Reson.*, 1990. **87**(2): p. 422-428.

180. Macura, S. and Ernst, R.R., Elucidation of Cross Relaxation in Liquids by Two-Dimensional Nmr-Spectroscopy. *Mol. Phys.*, 1980. **41**(1): p. 95-117.

181. States, D.J., Haberkorn, R.A., and Ruben, D.J., A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *J. Magn. Reson.*, 1982. **48**(2): p. 286-292.

182. Johnson, B.A., Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.*, 2004. **278**: p. 313-52.

183. Guntert, P., Mumenthaler, C., and Wuthrich, K., Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.*, 1997. **273**(1): p. 283-98.

184. Herrmann, T., Guntert, P., and Wuthrich, K., Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.*, 2002. **319**(1): p. 209-27.

185. Waldo, G.S., Standish, B.M., Berendzen, J., and Terwilliger, T.C., Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.*, 1999. **17**(7): p. 691-5.

186. McRee, D.E., XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, 1999. **125**(2-3): p. 156-65.

187. Gallivan, J.P. and Dougherty, D.A., Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.*, 1999. **96**(17): p. 9459-64.

188. Tatko, C.D. and Waters, M.L., The geometry and efficacy of cation-pi interactions in a diagonal position of a designed beta-hairpin. *Protein Sci.*, 2003. **12**(11): p. 2443-52.

189. Tatko, C.D. and Waters, M.L., Comparison of C-H...pi and hydrophobic interactions in a beta-hairpin peptide: impact on stability and specificity. *J. Am. Chem. Soc.*, 2004. **126**(7): p. 2028-34.

190. Hughes, R.M. and Waters, M.L., Influence of N-methylation on a cation-pi interaction produces a remarkably stable beta-hairpin peptide. *J. Am. Chem. Soc.*, 2005. **127**(18): p. 6518-9.

191. Hughes, R.M. and Waters, M.L., Effects of lysine acetylation in a beta-hairpin peptide: comparison of an amide-pi and a cation-pi interaction. *J. Am. Chem. Soc.*, 2006. **128**(41): p. 13586-91.

192. Wishart, D.S. and Case, D.A., Use of chemical shifts in macromolecular structure determination. *Methods Enzymol.*, 2001. **338**: p. 3-34.

193. Stanger, H.E., Syud, F.A., Espinosa, J.F., Giriat, I., Muir, T., and Gellman, S.H., Length-dependent stability and strand length limits in antiparallel beta -sheet secondary structure. *Proc. Natl. Acad. Sci. U.S.A.*, 2001. **98**(21): p. 12015-20.

194. Proba, K., Worn, A., Honegger, A., and Pluckthun, A., Antibody scFv fragments without disulfide bonds made by molecular evolution. *J. Mol. Biol.*, 1998. **275**(2): p. 245-53.

195. Ohage, E.C., Wirtz, P., Barnikow, J., and Steipe, B., Intrabody construction and expression. II. A synthetic catalytic Fv fragment. *J. Mol. Biol.*, 1999. **291**(5): p. 1129-34.

196. Bodenmuller, H., Schilling, E., Zachmann, B., and Schaller, H.C., The neuropeptide head activator loses its biological acitivity by dimerization. *Embo. J.*, 1986. **5**(8): p. 1825-9.

197. Lai, J.R. and Gellman, S.H., *Reinvestigation of the proposed folding and self-association of the Neuropeptide Head Activator. Protein Sci.*, 2003. **12**(3): p. 560-6.

198. Barbas III, C.F., and Wagner, J., Synthetic Human Antibodies: Selecting and Evolving Functional Proteins. *Methods*, 1995. **8**: p. 94-103.

199. Devlin, J.J., Panganiban, L.C., and Devlin, P.E., Random Peptide Libraries - a Source of Specific Protein-Binding Molecules. *Science*, 1990. **249**(4967): p. 404-406.

200. Giebel, L.B., Cass, R., Milligan, D.L., Young, D., Arze, R., and Johnson, C., Screening of cyclic peptide phage libraries identifies ligands that bind streptavidin with high affinities. *Biochemistry*, 1995. **34**(47): p. 15430-15435.

201. Weber, P.C., Pantoliano, M.W., and Thompson, L.D., Crystal structure and ligand-binding studies of a screened peptide complexed with streptavidin. *Biochemistry*, 1992. **31**(39): p. 9350-4.

202. Voet, D., Voet, J.G., and Pratt, C.W., Fundamentals of Biochemistry. 1999: Wiley.

203. Ho, K.L., Yusoff, K., Seow, H.F., and Tan, W.S., Selection of high affinity ligands to hepatitis B core antigen from a phage-displayed cyclic peptide library. *J. Med. Viro.*, 2002. **69**(1): p. 27-32.

204. Amin, A., Zaccardi, J., Mullen, S., Olland, S., Orlowski, M., Feld, B., Labonte, P., and Mak, P., Identification of constrained peptides that bind to and preferentially inhibit the activity of the hepatitis C viral RNA-dependent RNA polymerase. *Virology*, 2003. **313**(1): p. 158-69.

205. Lee, S.W., Mao, C., Flynn, C.E., and Belcher, A.M., Ordering of quantum dots using genetically engineered viruses. *Science*, 2002. **296**(5569): p. 892-5.

APPENDIX

XTZ1-PEPTIDE NMR AND STRUCTURE DATA

Table A1 Chemical shifts for **xtz1**-peptide

Residue	HN[a]	Hα	Hβ,β'	Others
Lys 1	exchange	4.064	1.957 1.896	Hγ,γ' 1.531; Hδ,δ' 1.714; Hε,ε' 3.059
Ala 2	exchange	4.473	1.476	
Trp 3	8.329 (8.2)	5.010	2.832	Hδ1 7.197; Hε1 10.218; Hε3 7.041; Hζ2 7.484; Hζ3 7.617; Hη2 7.256
Thr 4	9.091 (8.2)	4.880	4.131	Hγ 1.302
Trp 5	9.169	5.060	3.273 2.591	Hδ1 6.937; Hε1 9.451; Hε3 6.361; Hζ2 7.159; Hζ3 6.794; Hη2 7.107
Thr 6	9.503 (9)	4.849	3.980	Hγ 1.228
Trp 7	8.782 (5.5)	3.792	2.477 1.300	Hδ1 6.497; Hε1 9.976; Hε3 5.646; Hζ2 7.369; Hζ3 6.670; Hη2 7.022
Asn 8	7.315	4.850	3.110 2.389	Hδ,δ' 7.303, 7.265
Pro 9		3.571	2.391 1.959	Hγ,γ' 2.043; Hδ,δ' 3.598, 3.767
Ala 10	7.752 (6.2)	4.166	1.433	
Thr 11	6.838 (9.7)	4.336	4.193	Hγ 1.045
Gly 12	8.019	3.832 3.258		
Lys 13	6.711 (9.4)	4.668	1.740 1.801	Hγ,γ' 1.329, 1.472; Hδ,δ' 1.639; Hε,ε' 3.055
Trp 14	8.700 (8.3)	4.889	2.228 2.792	Hδ1 7.458; Hε1 10.140; Hε3 6.897; Hζ2 7.209; Hζ3 7.150; Hη2 7.129

Thr 15	8.960 (5.8)	4.657	4.134	Hγ 1.152
Trp 16	8.732 (7.2)	5.160	3.107 2.701	Hδ1 7.176; Hε1 9.797; Hε3 6.979; Hζ2 7.255; Hζ3 7.055; Hη2 7.146
Arg 17	9.120 (8.5)	4.744	1.919 1.741	Hγ,γ' 1.620; Hδ,δ' 3.186; Hε 7.167
Lys 18	8.475	3.722	1.257 0.854	Hγ,γ' 0.741, 0.650; Hδ,δ' 1.374, 1.313; Hε,ε' 2.697
Asn 19	8.439 (7.4)	4.712	2.839 2.652	Hδ,δ' not observed
Glu 20	8.056 (7.4)	4.124	2.068 1.881	Hγ,γ' 2.186
[a] Numbers in parenthesis are $J_{HN-H\alpha}$ in hertz (Hz).				

Table A2 Statistics for the 20 model NMR solution structures of **xtz1**-peptide

Number of residues	20
Average CYANA target function ($Å^2$)	0.22
Distance restraints (#)	
All	309
Short-range, $\|i - j\| \leq 1$	158
Medium-range, $1 < \|i - j\| < 5$	47
Long-range, $\|i - j\| \geq 5$	104
Distance Bounds (# restraints)	
- 2.99	12 (3.9 %)
3.00 - 3.99	94 (30.4 %)
4.00 - 4.99	140 (45.3 %)
5.00 - 5.99	63 (20.4 %)
6.00 -	0 (0 %)
RMSD	
All residues, backbone (Å)	0.27 +/- 0.09
All residues, heavy atoms (Å)	0.74 +/- 0.15
Ramachandran plot statistics	
Most favored regions	56.6 %
Additional allowed regions	43.4 %
Generously allowed regions	0.0 %
Disallowed regions	0.0 %